



OTEVŘENÁ DATA VE STÁTNÍ SPRÁVĚ

Diplomová práce

Studijní program: N6209 – Systémové inženýrství a informatika

Studijní obor: 6209T021 – Manažerská informatika

Autor práce: **Bc. Radek Štajner**

Vedoucí práce: doc. Ing. Klára Antlová, Ph.D.





TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Economics



OPEN DATA IN STATE ADMINISTRATION

Diploma thesis

Study programme: N6209 – System Engineering and Informatics

Study branch: 6209T021 – Managerial Informatics

Author: **Bc. Radek Štajner**

Supervisor: doc. Ing. Klára Antlová, Ph.D.



ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Radek Štajner**
Osobní číslo: **E12000030**
Studijní program: **N6209 Systémové inženýrství a informatika**
Studijní obor: **Manažerská informatika**
Název tématu: **Otevřená data ve státní správě**
Zadávající katedra: **Katedra informatiky**

Z á s a d y p r o v y p r a c o v á n í :

1. Charakteristika otevřených dat
3. Otevřená data v ČR a v zahraničí
4. Možnosti využití otevřených dat ve státní správě
5. Návrh standardizace datových setů

Rozsah grafických prací:

Rozsah pracovní zprávy:

65 normostran

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

GURIN, J. Open Data Now: The Secret to Hot Startups, Smart Investing, Sawy Marketing and Fast Innovation. 1st. ed. New York: McGraw-Hill, 2014. ISBN 0071829784.

BOČEK, J., J. MRÁČEK a J. MYNARZ. Otevřená data: Příležitost pro Českou republiku. 1. vyd. Praha: Nadace Open Society Fund, 2012. ISBN 8087725026.

MUSTAFA, M. Corporate Use of Open Government Data. 1st ed. Vienna: Technical University of Vienna, 2012. [vid. 2014-10-05]. Dostupné z: <http://cdn.open3.at/wp-content/uploads/2012/06/MM-OGD-JH-Final-v2.pdf> Elektronická databáze článků ProQuest (knihovna.tul.cz).

Vedoucí diplomové práce:

doc. Ing. Klára Antlová, Ph.D.

Katedra informatiky

Konzultant diplomové práce:

Viktor Ponomarev

ABRA, a. s. Praha

Datum zadání diplomové práce:

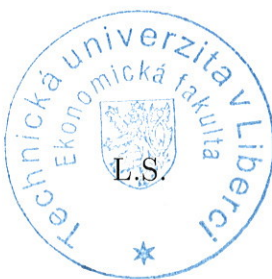
31. října 2014

Termín odevzdání diplomové práce:

7. května 2015



doc. Ing. Miroslav Žižka, Ph.D.
děkan



doc. Ing. Jan Skrbek, Dr.
vedoucí katedry

V Liberci dne 31. října 2014

Prohlášení

Byl jsem seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu TUL.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Diplomovou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé diplomové práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum:

Podpis:

Poděkování

Rád bych touto cestou poděkoval vedoucí diplomové práce paní doc. Ing. Kláře Antlové, Ph.D., za odborné vedení diplomové práce.

Panu Viktoru Ponomarevovi ze společnosti ABRA a.s. a panu Ing. Tomáši Kejzlarovi ze statutárního města Děčín děkuji za cenné rady z praxe, které byly nezbytné k napsání této diplomové práce.

Anotace

Tato diplomová práce se zabývá otevřenými daty v české veřejné správě. V teoretické části se zaměřuje na to, co otevřená data jsou, jaké jsou formáty a nástroje pro jejich zpracování. V následující části je provedena podrobná analýza situace otevřených dat v ČR a ve světě. Je zde rozebrána situace na jednotlivých úrovních státní správy a jsou uvedeny jednotlivé organizace či projekty zabývající se problematikou otevřených dat. V této kapitole je též poskytnut právní náhled na legislativu související s problematikou otevřených dat. Jedna kapitola je věnována problematice otevřených propojitelných dat. V další kapitole jsou popsány možnosti, potenciál a problémy otevřených dat ve státní správě. V oblastech, kde autor vidí nedostatky, se v závěrečné části snaží navrhnout jejich možná řešení. Autor v této kapitole navrhuje postup pro tvorbu univerzálních datasetů, které umožní efektivně porovnávat jednotlivá data mezi různými úřady či institucemi. Hlavním přínosem a výsledkem práce je pro čtenáře vytvoření uceleného přehledu o problematice otevřených dat ve státní správě.

Klíčová slova: Otevřená data, otevřená a propojitelná data, veřejná správa, vládní data, katalog otevřených dat, dataset, informace veřejného sektoru

Annotation

This diploma thesis deals with open data in Czech state administration. In theoretical part it covers what open data are, which formats and tools are there for their processing. In next part it analyzes current and past state of open data in Czech republic and the world. It also analyzes the situation in the various levels of state administration and are given examples of organizations or projects dealing with open data. This chapter also provide legal insight on the legislation related to the issue of open data. One chapter covers the issue of open linked data. In the next chapter are described possibilities, potential and problems of open data in state administration. In areas where the author sees shortcomings, he tries to suggest it's possible solutions. The author in this chapter suggests steps to creation of universal datasets which allows efficiently compare individual data between different agencies and institutions. The main contribution and goal of this thesis is the creation of integrated summary on the issue of open data in government.

Keywords: Open data, open linked data, state administration, government data, open data catalog, dataset, public sector information

Obsah

Anotace	7
Annotation	8
Seznam obrázků.....	11
Seznam tabulek.....	12
Seznam zkratk a značek:.....	13
Úvod	15
1 Otevřená data.....	17
1.1 Definice	17
1.1.1 Otevřená data.....	18
1.1.2 Propojitelná data.....	19
1.1.3 Data veřejné správy	19
1.2 Přínosy otevřených dat	21
1.3 Stupně otevřenosti	23
1.4 Formáty	24
1.5 Aplikace a nástroje pro práci.....	25
1.6 Rizika a problémy	26
2 Otevřená data v ČR a v zahraničí	27
2.1 Organizace a projekty v ČR	27
2.2 Katalog otevřených dat veřejné správy ČR.....	31
2.3 Celostátní úroveň.....	34
2.4 Regionální úroveň	35
2.4.1 Kraje	35
2.4.2 Liberecký kraj.....	36
2.4.3 Kraj Vysočina.....	37
2.4.4 Města	38
2.4.5 Město Děčín.....	39
2.5 Legislativa v ČR.....	40
2.5.1 Zveřejňovaná data.....	40
2.5.2 Nezveřejňovaná data.....	41
2.6 Zahraničí.....	42

2.6.1	Evropa.....	42
2.6.2	Svět	46
2.7	Organizace a projekty v zahraničí	49
2.8	Soukromý sektor.....	53
3	Propojitelná otevřená data	56
3.1	Definice LOGD	56
3.2	Výhody LOD.....	57
3.3	Aplikace a nástroje pro práci.....	57
3.4	LOD v ČR	58
3.5	LOD v zahraničí	59
4	Možnosti využití dat ve státní správě	62
4.1	Současná pozice české státní správy	62
4.2	Problémy otevřených dat v ČR	63
4.3	Potenciál otevřených dat ve státní správě.....	64
4.4	Kroky ke zlepšení využitelnosti otevřených dat v ČR	64
5	Návrh standardizace datových setů	66
5.1	Dataset.....	66
5.2	Vhodné oblasti pro datasety	66
5.3	Standardizace datových setů	67
5.4	Formát dat.....	69
5.5	Práce s RDF.....	70
5.5.1	Převod do RDF	70
5.5.2	Uložení RDF.....	72
5.5.3	Prohlížení RDF	74
5.6	Navrhovaný proces pro zavedení standardizace	76
6	Vyhodnocení návrhu	78
6.1	Nároky na implementaci	79
6.2	Výhody standardizace datasetů	79
	Závěr.....	80
	Seznam použité literatury	82
	Seznam příloh.....	86

Seznam obrázků

Obrázek 1. - Struktura veřejné správy ČR.....	20
Obrázek 2. - Otevřená propojitelná data veřejné správy	21
Obrázek 3. - Pětihvězdičkové schéma otevřenosti dat	23
Obrázek 4. - Portál veřejné správy ČR - portal.gov.cz.....	28
Obrázek 5. - Road Safety - aplikace pro zobrazení nejčastějších míst střetu se zvěří.....	30
Obrázek 6. - Katalog dat České republiky.....	33
Obrázek 7. - Webová aplikace Monitor.....	35
Obrázek 9. - Rozklikávací rozpočet Libereckého kraje	37
Obrázek 10. - Portál otevřených dat kraje Vysočina	38
Obrázek 11. Britský portál pro otevřená data - data.gov.uk.....	43
Obrázek 12. - Walkonomics - aplikace pro zobrazení průchodnosti ulic.....	44
Obrázek 13. - Rakouský portál pro otevřená data - data.gv.at	45
Obrázek 14. - Americký portál pro otevřená data - data.gov	48
Obrázek 15. - Global open data index	49
Obrázek 16. - Země zapojené do OGP	50
Obrázek 17. - Rozložení rozpočtu Horizon 2020	51
Obrázek 18. - Mapa propojení LOD.....	56
Obrázek 19. - Životní cyklus LOD	60
Obrázek 20. - Rakouský portál pro otevřená propojitelná data - lodpilot.at	61
Obrázek 21. - Open data index 2014 - pořadí zemí.....	62
Obrázek 22. - Vhodné oblasti pro zveřejnění dat	67
Obrázek 23. - OpenRefine - nahrání souboru s daty	71
Obrázek 24. - OpenRefine - tabulka.....	72
Obrázek 25. - Turtle syntaxe	73
Obrázek 26. - Apache Jena Fuseki - nahrání datasetu	74
Obrázek 27. - Apache Jena Fuseki - SPARQL dotazování	75

Seznam tabulek

Tab. 1 - Vhodnost jednotlivých formátů dat ke zveřejnění	24
Tab. 2 - Role při publikaci a katalogizaci otevřených dat	32
Tab. 3. - Srovnání grantů v Přerově a Opavě	68
Tab. 4 - Vyhodnocení návrhů	78

Seznam zkratek a značek:

*.ttl	Přípona názvu Turtle soubor
CIP	Competitiveness and Innovation Framework Programme - rámcový program pro konkurenceschopnost a inovace
CKAN	Comprehensive Knowledge Archive Network
CSV	Comma-separated values - hodnoty oddělené čárkami
ČOI	Česká obchodní inspekce
ČSÚ	Český statistický úřad
ČTÚ	Český telekomunikační úřad
DOC	Formát souboru programu Microsoft Word
EVVO	Enviromentální vzdělávání, výchova a osvěta
FO	Fyzická osoba
GPS	Global Positioning System - globální polohovací systém
HTTP	Hypertext Transfer Protocol - internetový protokol
JSON	JavaScript Object Notation - JavaScriptový objektový zápis
LOD	Linked Open Data - propojitelná otevřená data
LOGD	Linked Open Government Data - propojitelná otevřená data veřejné správy
MGI	McKinsey Global Institute
MVČR	Ministerstvo vnitra České republiky
OGP	Open Government Partnership - partnerství pro otevřené vládnutí
OKFN	Open Knowledge Foundation
OSF	Open Society Fund
PDF	Portable Document Format - přenosný formát dokumentů
PO	Právnícká osoba
PSI	Public sector information - informace veřejného sektoru
RDF	Resource Description Framework - systém popisu zdrojů
RTI	Right To Information - právo na informace
SPARQL	Dotazovací jazyk pro RDF databáze
SQL	Structured Query Language - strukturovaný dotazovací jazyk
SW	Software
TXT	Textový formát souborů

UK	Univerzita Karlova
URI	Unified Resource Identifier
VS	Veřejná správa
VŠE	Vysoká škola ekonomická
XLS	Formát souborů aplikace MS Excel
XML	Extensible Markup Language - rozšiřitelný značkovací jazyk
ŽP	Životní prostředí

Úvod

Množství a objemy dat vyprodukovaných v ekonomikách 21. století neustále rostou. Společně s tím roste i poptávka po těchto datech. Státní správa není výjimkou, sbírá obrovské množství informací a spousta z těchto informací je ze své podstaty veřejná. Díky moderním technologiím lze tato data nejen sbírat, ale je možné, aby byla veřejně komukoliv a kdykoliv k dispozici. Taková data je poté možné nazvat jako data otevřená.

Otevřená data jsou dnes velký fenomén. Přitom ještě před pár lety se této problematice nevěnovala téměř žádná pozornost. Otevřená data se začínají prosazovat v různých oblastech veřejného i soukromého sektoru. Tato diplomová práce je zaměřena především na otevřená data v oblasti veřejné správy.

V první kapitole je rozebrána podstata otevřených dat, k čemu slouží a jaké jsou jejich výhody. Jsou popsány jednotlivé formáty a aplikace pro práci s otevřenými daty. V závěru kapitoly jsou popsána i některá rizika a problémy s otevřenými daty spojená.

Druhá kapitola přináší ucelenou analýzu otevřených dat v České republice i v zahraničí. Nejdříve je popsána situace otevřených dat v ČR. Podrobně jsou rozebrány jednotlivé úrovně státní správy. Následuje analýza jednotlivých organizací a projektů. Jsou uvedeny příklady otevřených měst a krajů. Též je popsána příslušná legislativa s problematikou otevřených dat související. V závěru kapitoly je provedena stejná analýza jako v případě ČR i pro otevřená data v zahraničí.

Ve třetí kapitole jsou podrobněji rozebrána otevřená propojitelná data. Právě tento druh otevřených dat je označován jako cíl, kam by měla směřovat snaha všech, co mají zájem na otevírání dat. Stejně jako v případě otevřených dat je provedena analýza stavu otevřených propojitelných dat v ČR i ve světě.

V následující kapitole přichází přehled možností a potenciálu otevřených dat v českém veřejném sektoru. Na základě definovaných problémů jsou uvedeny kroky, které by se měly podniknout ke zlepšení stavu otevřených dat v ČR.

V předposlední kapitole přichází autor s návrhem vhodných metod a postupů pro standardizování datasetů jako možnosti k efektivnějšímu využití informací, které již mnoho institucí samo dobrovolně zveřejňuje, ale díky rozdílným přístupům v jejich zveřejnění dochází ke ztrátě potenciálu takto zveřejněných informací.

V poslední kapitole je stručné zhodnocení uvedených návrhů včetně posouzení reálnosti jejich uvedení do praxe. Jsou zde uvedeny výhody, nevýhody, finanční a technická náročnost těchto návrhů.

1 Otevřená data

Nejdříve by bylo vhodné si říci, co to vlastně otevřená data jsou. Celá myšlenka otevřených dat vychází z principů RTI (Right To Information – právo na informace) hnutí, které historicky navazuje na právo na svobodu projevu. Principy těchto práv byly stanoveny na konci 18. století v době Velké francouzské revoluce a společně s dalšími právy byly souhrnně vydány jako Deklarace práv člověka a občana. Tato deklarace vedla k vytvoření první francouzské ústavy a později i ke změně celého evropského vnímání těchto práv. Tato práva patří mezi základní principy demokratických režimů a jsou nezbytnou podmínkou k efektivnímu zavedení otevřených dat v dané zemi. Otevřená data samotná jsou v nejzákladnější podobě veřejně dostupné informace. Jedná se tradičně o data, která mají pro čtenáře těchto dat nějaký význam a která by bez zveřejnění vlastním bylo jinak komplikované získat. Mezi nejznámější příklady otevřených dat patří například jízdní řády, aktuálně rozestavěné silnice, informace o příjmech a výdajích státu či různé statistické údaje o obyvatelstvu.

1.1 Definice

Je potřeba získat přehled o několika podobných termínech, ve kterých by jinak bylo možné se snadno ztratit. V této diplomové práci se většinou mluví o otevřených datech, tato data mohou být též součástí dat veřejné správy a zároveň se může jednat o data propojitelná.

V průběhu celé práce se též objevují pojmy data, informace, případně znalosti. Tyto pojmy v běžné mluvě často splývají, přesto je nutné je z hlediska informačních technologií rozlišovat.

Pojem data se využívá pro čísla, text, zvuk, obraz, případně jiné vjemy reprezentované v podobě vhodné ke zpracování počítačem. Data jsou surovinou, ze které je možné získat informace. Informace jsou data v kontextu, použitelná a srozumitelná. Informace jsou podmnožinou pojmu znalosti. Znalost je možné chápat jako souvislosti plynoucí z provázané struktury informací. Znalosti jsou individuální – ze stejných informací mohou vzniknout různé znalosti. [1]

1.1.1 Otevřená data

Jednoduchá definice říká: „*Otevřená data jsou definována jako uložená data, která mohou být přístupná veřejnosti bez jakéhokoliv omezení užití či distribuce.*“ [2 s. 1]

Jsou to tedy například: „*data, která státní instituce vystavily na internet a daly volně k dispozici. Tak, aby je mohl využívat kdokoli: prohlížet si je, propojovat s daty z jiných zdrojů, analyzovat, vytvářet z nich zajímavé mapy a vizualizace nebo vyvíjet užitečné mobilní aplikace. Neziskově i komerčně.*“ [3 s. 4]

Dle MVČR otevřená data musí být:

- úplná,
- snadno dostupná,
- strojově čitelná,
- používající standardy s volně dostupnou specifikací,
- zpřístupněna za jasně definovaných podmínek užití dat s minimem omezení,
- dostupná uživatelům při vynaložení minima možných nákladů. [4]

Pravidla pro užití otevřených dat:

- neomezují jejich uživatele ve způsobu použití dat,
- opravňují uživatele k jejich dalšímu šíření,
- musí být uveden autor dat (i při dalším šíření),
- při dalším šíření musí i ostatní uživatelé mít stejná oprávnění s daty
nakládat - během šíření dat nesmí dojít např. k omezení jejich využití pouze pro
nekomerční účely. [4]

Právní definice toho, co to otevřená data jsou, v současné době neexistuje. Není stanoven ani žádný oficiální standard. Je tak především z důvodu, že celá problematika zavádění otevřených dat se dá označit za komunitní snahu určité skupiny lidí.

1.1.2 Propojitelná data

Pojem propojitelná data popisuje situaci, kdy jsou mezi sebou určitým způsobem propojena data z více zdrojů a mohou se tak stát více užitečnými.

Vynálezce webu Tim Bernes-Lee v roce 2006 uvedl 4 pravidla pro propojitelná data:

1. Používat URI k pojmenování dat.
2. Používat HTTP URI, aby lidé mohli tato jména vyhledat.
3. Pokud si uživatel vyhledá URI, je nutné mu poskytnout užitečné informace za použití vhodného formátu dat (RDF, SPARQL).
4. Poskytnout propojení k dalším URI, pomocí nichž může uživatel získat více informací.

Propojením otevřených a propojitelných dat vznikají tzv. otevřená propojitelná data, neboli LOD (Linked Open Data), kterými se více zabývá kapitola 3.

1.1.3 Data veřejné správy

Data veřejné správy jsou veškerá data, jejichž vlastníkem jsou instituce veřejné správy. Vlastník dat je tradičně instituce, PO nebo FO, u které data vznikla. Veřejná správa je soustava úřadů vykonávající služby veřejného sektoru (kromě úřadů vykonávají služby veřejného sektoru i nemocnice, školy a další). Veřejný sektor je jednou ze součástí systému ekonomiky, tou druhou je sektor soukromý, někdy se uvádí ještě třetí sektor neziskový. Veřejná správa je rozdělena na dva hlavní aktéry – státní správa a samospráva. Oba aktéři by se měli rozhodovat na základě veřejného zájmu neboli potřeb společnosti. [5]

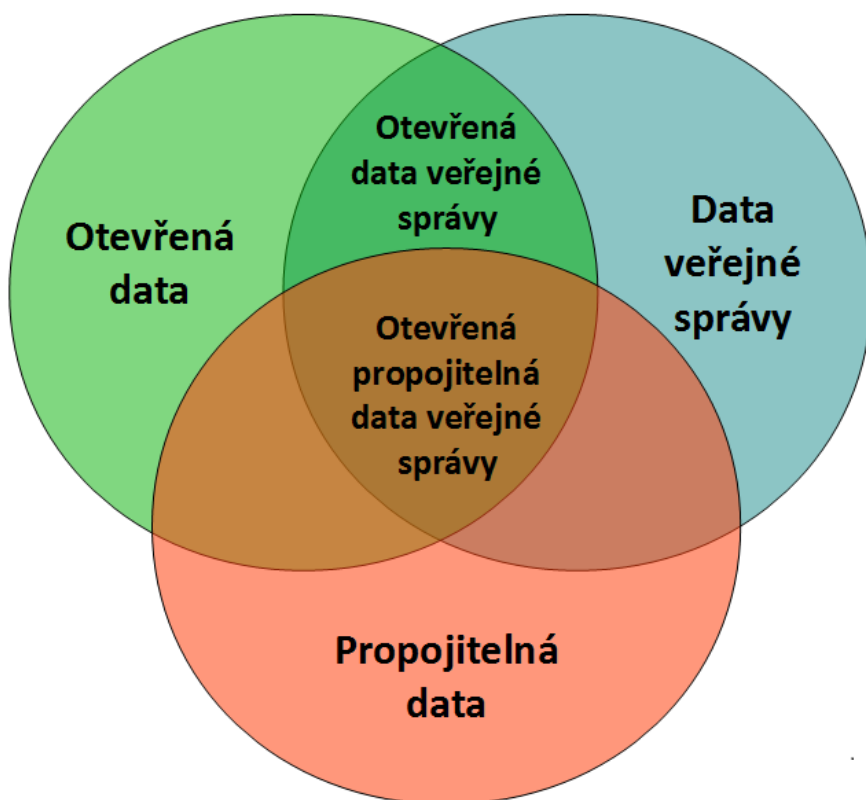
Struktura veřejné správy ČR



Obrázek 1. - Struktura veřejné správy ČR

Zdroj: http://www.garep.cz/wp-content/gallery/vzdelavani/struktura_vs.jpg

Obrázek 2 ukazuje propojení pojmů otevřená data, propojitelná data a data veřejné správy. Vznikají tak otevřená propojitelná data veřejné správy neboli LOGD (Linked Open Government Data).



Obrázek 2. - Otevřená propojitelná data veřejné správy

Zdroj: Kučera J. a D. Chlapek, Přínosy a rizika otevřených a propojitelných dat

Otevřená data jsou z pohledu veřejné správy součástí služeb souhrnně označovaných jako eGovernment.

„eGovernment je využívání informačních technologií veřejnými institucemi pro zajištění výměny informací s občany, soukromými organizacemi a jinými veřejnými institucemi za účelem zvyšování efektivity vnitřního fungování a poskytování rychlých, dostupných a kvalitních informačních služeb.“ [6 s. 7]

1.2 Přínosy otevřených dat

President OKFN Rufus Pollock kdysi řekl: „*Nejzajímavější způsob využití vašich dat vymyslí někdo jiný.*“ Právě díky tomuto principu je možné z dat vytěžit více, než by se na první pohled mohlo zdát.

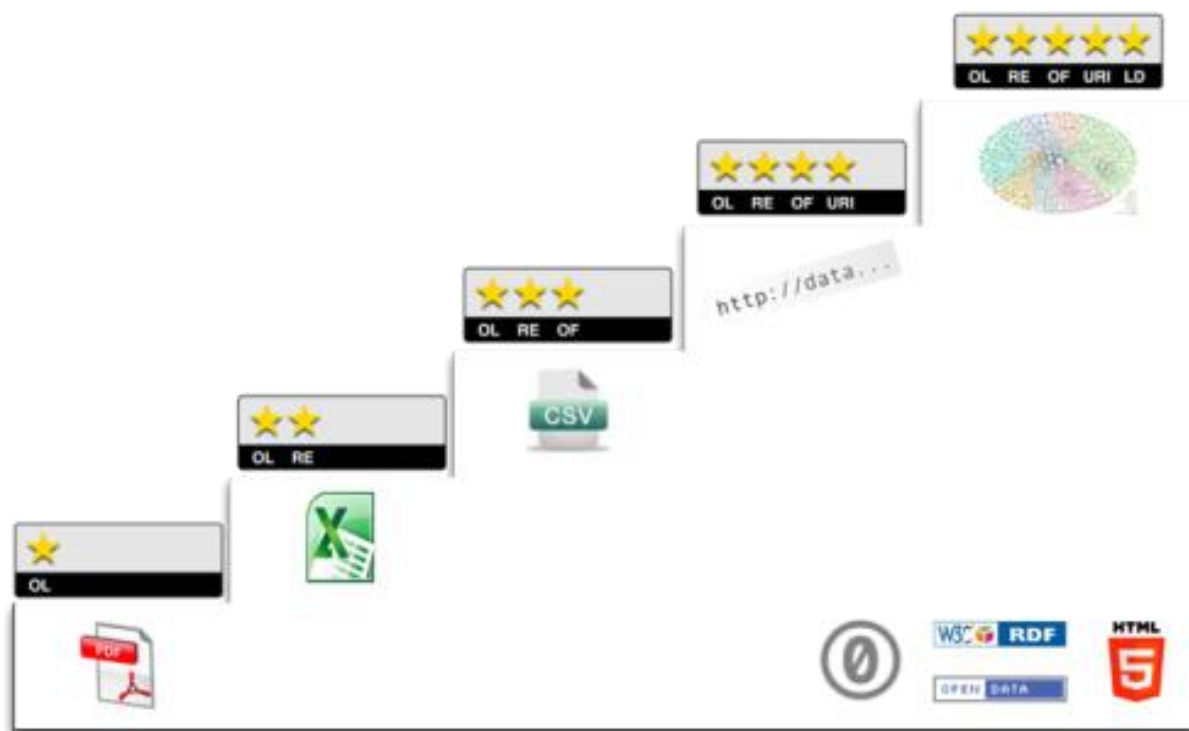
Hlavním důvodem zavádění principů otevřených dat jsou jejich přínosy, mezi ně patří:

- **Zvýšení efektivity:** uvolnění dat znamená možnost je sdílet a analyzovat a tím zvýšit jejich hodnotu.
- **Podpora ekonomiky:** data jsou zdrojem inovací, podnikatelských příležitostí a pracovních nabídek – lze je využít třeba v dopravě, logistice, zdravotnictví či bankovníctví. Firmy pracují s daty jako se surovinou, vytváří nad nimi aplikace, které generují přidanou hodnotu a zisk.
- **Transparentnost,** zefektivnění, zlepšení procesů a kontrola veřejné správy: zveřejněná data umožňují kontrolu, jak se hospodaří s daněmi nebo jaké jsou náklady organizací, které podporujeme.
- **Zlepšení služeb veřejné správy** a zlepšení jejího vnímání.
- **Zapojení občanů do rozhodování:** občané se mohou díky datům a analýzám kvalifikovaněji podílet na fungování státu.
- **Datová žurnalistika:** otevřená data jsou nezastupitelným zdrojem informací pro novináře.
- **Posílení demokracie:** otevřená data zefektivňují fungování demokracie. [4]

1.3 Stupně otevřenosti

Konkrétní datový soubor může být otevřený v různé míře. Tim Berners-Lee vytvořil takzvané pětihvězdičkové schéma otevřenosti dat. Na Obrázku 3 je toto schéma znázorněno.

1. Data jsou otevřená. (např. PDF)
2. Data jsou strojově čitelná. (např. *.xls)
3. Data jsou uložena ve formátech nezávislých na platformě. (např. CSV)
4. Datům jsou přiřazeny URI. (např. RDF)
5. Data jsou propojena s jinými daty (LOD).



Obrázek 3. - Pětihvězdičkové schéma otevřenosti dat

Zdroj: Hausenblas M., 5 star Open Data. Dostupné z: <http://5stardata.info/>

1.4 Formáty

Ke zveřejnění dat se využívá celá řada různých formátů. Následující tabulka popisuje jejich vhodnost.

Tab. 1 - Vhodnost jednotlivých formátů dat ke zveřejnění

Formát	Nezávislost na aplikaci	Zápis ve strukturované podobě	Popis struktury dat	Popis sémantiky dat	Vytváření propojení	Vhodnost použití 1=nejlepší
PDF	Ne	Ne	Ne	Ne	Ne	5
DOC(X)	Ne	Ne	Ne	Ne	Ne	5
TXT	Ano	Ne	Ne	Ne	Ne	5
HTML	Ano	Částečně	Ne	Ne	Ne	4
XLS(X)	Ne	Částečně	Ne	Ne	Ne	4
CSV	Ano	Ano	Částečně	Ne	Ne	3
JSON	Ano	Ano	Částečně	Ne	Ne	3
XML	Ano	Ano	Ano	Ne	Ne	2
OData	Ano	Ano	Ano	Částečně	Částečně	2
RDF	Ano	Ano	Ano	Ano	Ano	1

Zdroj: Chlapek D., J. Kučera a M. Nečaský, Metodika publikace otevřených dat veřejné správy ČR, vlastní zpracování

Nezávislost na aplikaci umožňuje vznik nových aplikací, které mohou následně pracovat s daným formátem, ve kterém jsou otevřená data k dispozici. Formát, který toto umožňuje, bývá označován jako otevřený formát/standard. Digital Standards Organization definuje otevřený standard jako:

- Svobodný a otevřený standard je nezávislý na výrobci ve všech fázích svého životního cyklu. To umožňuje volně využívat, vylepšit a rozšířit standard v průběhu času.
- Standard je přijat a bude udržován pomocí neziskové organizace. Jeho další rozvoj je tvořen na základě otevřeného rozhodovacího postupu a je k dispozici všem zainteresovaným.
- Standard byl publikován a dokument se specifikací standardu je volně k dispozici. Komukoliv musí být umožněno tento dokument kopírovat, distribuovat a volně používat.
- Patenty případně uvalené na (části) standardu jsou neodvolatelně k dispozici bezúplatně.
- Nejsou žádná omezení týkající se opětovného použití standardu. [7]

1.5 Aplikace a nástroje pro práci

Nástrojů pro práci s otevřenými daty je velké množství. Může se jednat o jakékoliv formáty souborů umožňující zveřejnění dat. Jak již bylo zmíněno dříve, zvolený formát může též určit, o jaký stupeň otevřenosti dat se jedná. Zvolená aplikace a zvolený formát spolu logicky souvisí, přičemž čím univerzálněji použitelný formát zvolíme, tím vyšší stupeň otevřenosti dat dostaneme. Může se jednat například o již klasické aplikace typu Microsoft Word či Microsoft Excel. Formáty těchto aplikací ovšem nepřinášejí dostatečný stupeň otevřenosti. Vyšší stupeň otevřenosti přinášejí formáty postavené na XML, OData nebo RDF. Tyto formáty mají otevřený standard a existuje pro ně tedy nepřeberné množství aplikací a nástrojů. Namátkou např.: Open Refine, Anzo Express, Triplify nebo Janitor.

Další možností pro práci s otevřenými daty jsou webové aplikace. Všechny tyto nástroje mají za úkol data nějakým způsobem zpracovat a povětšinou se jedná o aplikace, které nebyly primárně vytvořené pro práci s otevřenými daty. Vývoj SW nástrojů pro oblast otevřených dat řeší též několik projektů (viz kap. 2.7 Organizace a projekty v zahraničí).

Kromě těchto nástrojů vzniká samozřejmě i množství aplikací, které jsou cíleně vytvořené pouze pro otevřená data. Většinou přímo pro konkrétní oblast. Některé z těchto aplikací jsou uvedeny v příloze B.

1.6 Rizika a problémy

Otevřená data s sebou nepřinášejí jen výhody, ale v některých případech mohou přinést i problémy. Tyto problémy mohou vzniknout před zveřejněním, během zveřejňování i po zveřejnění. Noor Huijboom a Tijs van den Broek je shrnují takto:

- uzavřená vládní kultura,
- zveřejnění v rozporu se zákonem,
- právní ochrana soukromí a osobních údajů,
- riziko porušení obchodního tajemství,
- omezená/nízká kvalita dat,
- omezená/nízká přívětivost k uživateli či informační přehlcení,
- nízká úroveň standardizace otevřených dat,
- bezpečnostní hrozby,
- existující modely plateb za informace,
- nejistý ekonomický dopad,
- digitální rozdílnost,
- absence konzumentů dat,
- překrývání dat a síťové přetížení. [8], [9]

2 Otevřená data v ČR a v zahraničí

Vzhledem k poměrně mladému tématu otevřených dat a velké informační roztržitosti v této oblasti nemusí být úplně jednoduché pochopit co z informací, které má člověk možnost si přečíst jsou vlastně otevřená data. Tato kapitola přináší analýzu situace otevřených dat v ČR i ve světě, včetně popisu konkrétních organizací a projektů.

2.1 Organizace a projekty v ČR

V ČR v posledních letech vzniklo několik organizací, které se buď otevřenými daty přímo zabývají či se určitým způsobem snaží myšlenku otevřených dat podpořit. V roce 2014 mnoho českých úřadů, krajů a měst začalo zveřejňovat svá data na internetu. Otevřená data tak i v ČR získávají stále větší význam. Česká republika se v oblasti otevřených dat snaží inspirovat především od okolních zemí (Německo, Rakousko). Ve většině případů se nejedná o státní instituce, ale o komunitní organizace či projekty lidí, kteří se snaží o zavedení myšlenek otevřených dat v ČR.

Fond Otakara Motejla

FOM navazuje na myšlenky 1. českého ombudsmana (veřejný ochránce práv) Otakara Motejla. Tento fond byl založen za pomoci Nadace Open Society Fund Praha a jeho hlavním cílem je efektivní veřejná správa. Tohoto cíle chce mimo jiné dosáhnout prostřednictvím otevřených dat. Spolupracuje s dalšími nevládními organizacemi a vytváří vlastní projekty pro podporu otevřených dat. Jedním z hlavních úspěchů fondu je vznik projektu Rekonstrukce státu, který má jako hlavní cíl prosazení 9 klíčových zákonů pro zamezení korupce. Tento projekt mimo jiné sleduje i to, zda poslanci plní své sliby, týkající se prosazení těchto zákonů. Dalším zajímavým projektem je projekt Náš stát, který je určitým webovým rozcestníkem a přináší odkazy na projekty či aplikace využívající otevřených dat. Týkají se např. oblastí rozpočtů, životního prostředí nebo zákonů.

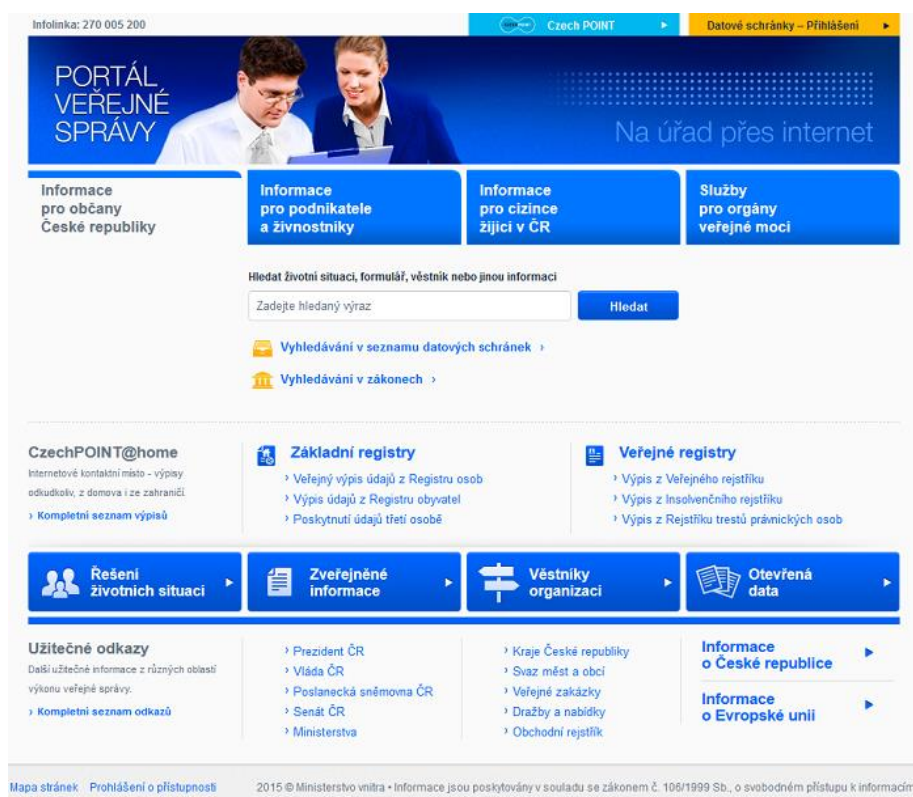
V roce 2013 se FOM podílel na pořádání soutěže o nejlepší české aplikace pro otevřená data. V soutěži bylo oceněno 25 aplikací využívající či poskytující otevřená data. Druhý ročník proběhl v roce 2014 a zúčastnilo se ho 24 aplikací.

opendata.cz

Iniciativa zabývající se problematikou otevřených dat. Jejím hlavním cílem je vybudování otevřené datové infrastruktury. Součástí tohoto plánu je i vybudování Katalogu dat ČR. Tvoří ho hlavně pedagogové a studenti Matematicko-fyzikální fakulty UK a Fakulty informatiky a statistiky VŠE. Po kliknutí na odkaz je možné přejít na doménu linked.opendata.cz, v dubnu 2015 se zde nacházelo 13 datasetů, např. kontroly ČOI nebo registry smluv.

Portál veřejné správy

Dalším místem, kde je možné nalézt otevřená data je web www.portal.gov.cz, pomocí něhož může občan komunikovat s úřady. Nejedná se o portál pouze pro otevřená data, ale pro různé informace v rámci eGovernmentu. Nabízí přístup k datovým schránkám, formulářům nebo veřejným registrům. Taktéž poskytuje užitečné informace pro cizince, prohledávání zákonů či například informace k řešení různých životních situací. Mimo jiné zde mohou instituce zveřejňovat své smlouvy a další dokumenty.



Obrázek 4. - Portál veřejné správy ČR - portal.gov.cz

Zdroj: portal.gov.cz/portal/obcan

Informace pro všechny

Iniciativa, která se pomocí webu www.infoprovsechny.cz snaží pomoci lidem s jejich dotazy na státní instituce. Pro běžného občana může být často komplikované vyznat se v rozsáhlé struktuře státních institucí. Na webu uživatel nalezne potřebné rady jak a komu svůj dotaz vznést. Zároveň je zde možnost vyhledávat v již vznesených dotazech a ušetřit si tak čas se vznášením vlastního dotazu. V březnu 2015 zde bylo evidováno přibližně 1600 dotazů. Tento projekt využívá zákonu číslo 106/1999 Sb., o svobodném přístupu k informacím.

Otevřená společnost

Skupina odborníků zabývající se například problematikou diskriminace, důvěryhodnosti policie, otevřeností veřejného prostoru a právem na informace. V těchto oblastech aktivně hledá nedostatky a problémy a snaží se navrhnout jejich možná řešení. Zároveň nabízí možnost školení, auditů či seminářů pro úřady, firmy i veřejnost. Mimo jiných je členem této neziskové organizace i Oldřich Kužílek, spoluautor Zákona číslo 106/1999 Sb., o svobodném přístupu k informacím. Oldřich Kužílek zároveň za projekt Otevřené společnosti dohlíží na jeho připravovanou novelu v roce 2015.

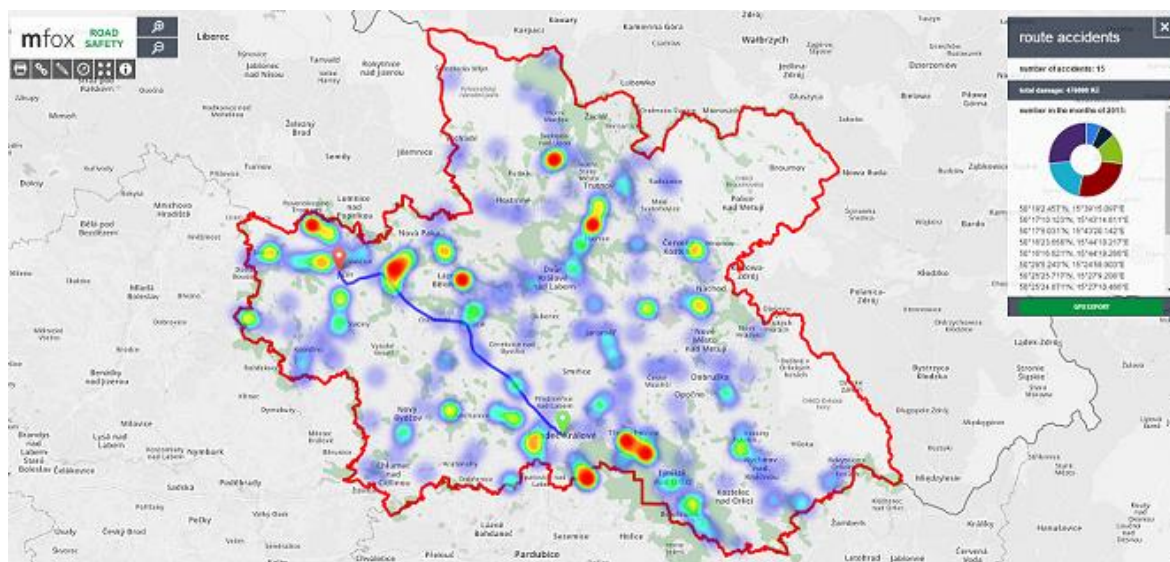
Nadace Open Society Fund Praha

Nadace prosazuje a financuje jak systémové tak i lokální projekty. Zároveň je zprostředkovatelem části Norských fondů a spravovatelem Nadačního fondu Hyundai. Hlavním dárce je George Soros a jeho síť Open Society Foundations. Mezi další dárce patří např. Google, Americká ambasáda či Britské velvyslanectví v Praze. Nadace OSF již podpořila více než 9300 projektů částkou přes 1,24 miliardy korun. Mezi hlavními podporovanými je i Fond Otakara Motejla.

mapazakazek.cz

Jedná se o projekt z konkrétní oblasti otevřených dat. Je postavený na Google mapách, kde jsou veřejné zakázky prezentovány jako body na mapě České republiky. Pomocí filtrů je pak možné upravit, které zakázky chce uživatel zobrazit. Na rozdíl od jednoduchých datasetů umožňuje data prezentovat v přehledné a srozumitelné formě.

Na podobném principu fungují i další projekty, kdy jsou otevřená data využívána společně s mapovým podkladem, například mapa dětských hřišť, mapa veřejných WC, mapa nejčastějších míst střetu aut se zvěří, cykloplánovač a mnoho dalších.



Obrázek 5. - Road Safety - aplikace pro zobrazení nejčastějších míst střetu se zvěří

Zdroj: <http://demo.mfox.cz/nav14/>

Další projekty či organizace podporující myšlenku otevřených dat:

Nadace Karla Janečka, Velvyslanectví USA, EEA Grants, kohovolit.eu, nadace NROS a další.

Úřad pro ochranu osobních údajů

V jistém slova smyslu se problematikou otevřených dat zabývá i Úřad pro ochranu osobních údajů, který vede dozor nad dodržováním zákonem stanovených povinností při zpracování osobních údajů. Nicméně se dá říci, že tento úřad stojí spíše na druhé straně bariéry a občas se v zájmu ochrany osobních údajů dostává do konfliktu s myšlenkou otevřených dat. Je ovšem nutné upozornit, že i dle definice Tim Bernse-Leeho nesmí být otevřená data osobního charakteru a zasahovat do soukromí lidí.

2.2 Katalog otevřených dat veřejné správy ČR

Důležitou roli v oblasti otevřených dat hraje katalog otevřených dat. Jedná se o jednotné místo pro vyhledávání otevřených dat veřejné správy a pro získávání údajů o těchto datech (tj. získávání metadat). Datový katalog neslouží primárně k ukládání samotných otevřených dat, ale k ukládání záznamů o těchto datech. Na základě těchto záznamů jsou uživatelé Datového katalogu schopni vyhledat data, která potřebují z jednoho místa bez nutnosti prohledávat izolované webové portály jednotlivých orgánů veřejné správy. Záznamy Datového katalogu také obsahují popis dat a tím usnadňují uživatelům práci s těmito daty. [10]

„Nalezitelnost a přehlednost jsou po přístupnosti na webu klíčové požadavky na otevřená data. Pokud data uživatel nenajde a nenajdou je ani webové vyhledávače, pak data jakoby neexistovala. Proto vznikají datové katalogy, které pomáhají lidem data najít a uspořádat“
[11 s. 1]

Metodika publikace otevřených dat veřejné správy ČR stanoví role při publikaci a katalogizaci otevřených dat:

Tab. 2 - Role při publikaci a katalogizaci otevřených dat

Role	Popis role	Odpovědnost
Poskytovatel dat	Jakýkoliv orgán veřejné správy, který má v rámci svých kompetencí právo či povinnost data zveřejňovat.	<ul style="list-style-type: none"> • zveřejnění dat ve formě otevřených dat, • určování podmínek, za jakých jsou data zveřejněna ve formě otevřených dat, • pověření kurátora správou dat.
Kurátor dat	Zajišťuje zveřejnění otevřených dat a údržbu záznamů o datech v Datovém katalogu.	<ul style="list-style-type: none"> • zveřejnění otevřených dat, • vytvoření, aktualizaci a zneplatnění záznamu o poskytnutých otevřených datech v Datovém katalogu, včetně zařazení záznamů do příslušných klasifikačních struktur.
Správce Datového katalogu	Subjekt plnící roli správce Datového katalogu může plnit i roli provozovatele Datového katalogu.	<ul style="list-style-type: none"> • zajištění výběru nástroje, implementace vybraného nástroje, • definici parametrů, sběr podnětů a zajištění provozu Datového katalogu, • určování redaktorů.
Provozovatel Datového katalogu	Zajištění běžného chodu, bezpečnosti a údržby Datového katalogu a potřebného softwarového a hardwarového vybavení.	<ul style="list-style-type: none"> • zajištění realizace a testování úprav parametrů a konfigurace prostředí pro provoz Datového katalogu podle pokynů správce Datového katalogu, • provoz Datového katalogu v rámci dohodnuté úrovně poskytovaných služeb, • zálohování a archivaci Datového katalogu.
Redaktor	Obsahová kontrola vložených záznamů o otevřených datech VS ČR.	<ul style="list-style-type: none"> • ověření korektnosti a klasifikace záznamu v Datovém katalogu, • komunikaci s poskytovateli dat ohledně kvality záznamu v Datovém katalogu, • označení záznamu ke zveřejnění.
Koncový uživatel	Vyhledává záznamy o otevřených datech VS ČR v Datovém katalogu,	<ul style="list-style-type: none"> • posílá správci Datového katalogu podněty na vytvoření nových záznamů, resp. úpravu stávajících záznamů v Datovém katalogu, • využívá data v souladu s podmínkami jejich užití.

Zdroj: Chlapek D., J. Kučera a M. Nečaský, Metodika publikace otevřených dat veřejné správy ČR, vlastní zpracování

Tuto metodiku vytvořil v roce 2012 Dušan Chlapek společně s kolegy a kromě rolí pro katalogizaci a údržbu otevřených dat stanoví i procesy a požadavky na vhodný katalogizační software.

Oficiální katalog otevřených dat veřejné správy ČR v současné době neexistuje. I proto došlo k vytvoření neoficiálního katalogu nazvaného Katalog dat České republiky.

Katalog dat České republiky

Jedná se neoficiální veřejný datový katalog. O založení katalogu se postarala iniciativa OpenData.cz. Katalog neobsahuje samotná data, ale pouze jejich popis. Je postaven na nástroji CKAN, vytvořeném organizací Open Knowledge Foundation. Tento nástroj se stal víceméně standardem pro katalogy po celém světě, využívá ho například britský datový katalog na portále data.gov.uk.

V současné době Katalog dat České republiky obsahuje přibližně 200 datasetů.



CKAN Czech Republic — Přihlásit Zaregistrovat

Přidat dataset Vyhledat Skupiny O CKAN Nalít datasety

Vítejte na CKAN Czech Republic!

Vyhledat data

Najít datasety

CKAN Czech Republic obsahuje **182 datasetů** to můžete prohlížet, poznávat a stahovat.

Sdílení dat

Vytvořte vlastní datasety a podělte se o ně s ostatními a najděte další, kteří se zajímají o vaše data.

Vytvoření datasetu »

Spolupracovat

Více se o práci s otevřenými daty můžete dozvědět na těchto zdrojích:

- [GetTheData.org](#)
- [DataPatterns.org](#)
- [Open Data Handbook](#)

Kdo sem přispívá?

Czech Republic Czech Republic má 145 datasetů. Fond Otakara Motejla Datasety watchdogových a protikorupčních organizací, které spolupracují s Fondem Otakara Motejla. Fond Otakara Motejla má 13 datasetů.	UEP - survey UEP - survey má 140 datasetů. Praha 8 Datasety městské části Praha 8 Praha 8 má 3 datasetů.	OpenData.cz Datasety zpracované OpenData.cz OpenData.cz má 18 datasetů.
--	--	--

Obrázek 6. - Katalog dat České republiky

Zdroj: <http://cz.ckan.net/>

2.3 Celostátní úroveň

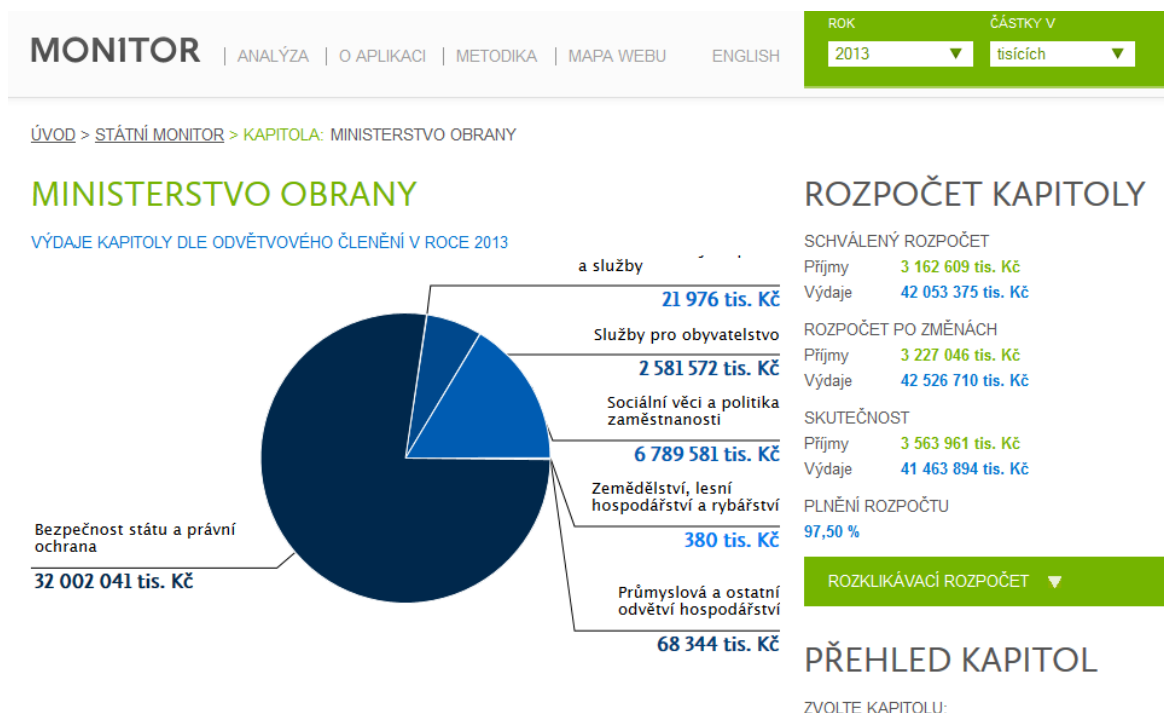
Různé veřejné instituce přistupují k otevřeným datům rozdílně. Proto je nutné odlišit celostátní úroveň veřejné správy od regionální. Nejdříve bude popsána celostátní úroveň a následně regionální.

ČR přistoupila k iniciativě Partnerství pro otevřené vládnutí usnesením vlády ze dne 14. září 2011. Následně byl 4. dubna 2012 schválen povinný „Akční plán České republiky Partnerství pro otevřené vládnutí“. Akční plán je v tomto případě souhrn návrhů, postupů a cílů vedoucích ke zlepšení situace v oblasti otevřených dat. V říjnu 2014 byl vydán aktualizovaný „Akční plán České republiky Partnerství pro otevřené vládnutí na období let 2014 až 2016“. Česká republika se v něm zavázala k:

- Přijetí nového zákona o státní službě zajišťujícího odpolitizování, profesionalizaci a stabilizaci státní správy a jeho implementace do praxe.
- Zefektivnění systému svobodného přístupu k informacím.
- Zpřístupnění dat a informací.

Mezi další strategické dokumenty patří: Strategický rámec rozvoje eGovernmentu 2014+, Digitální Česko v 2.0 – Cesta k digitální ekonomice, Strategie vlády v boji s korupcí na období let 2015-2017 a další.

Jako jeden z prvních datasetů byly během roku 2011 zpřístupněny prostřednictvím webu Českého statistického úřadu výsledky voleb. ČSÚ zveřejňuje i některé další údaje, které je možné označit za otevřená data, například výsledky sčítání lidu, domů a bytů. Kromě ČSÚ zveřejňuje otevřená data také ČOI, ČTÚ a další. Ministerstvo financí provozuje na webové adrese monitor.statnipokladna.cz aplikaci Monitor, která umožňuje občanům přístup k rozpočtovým a účetním informacím veřejné správy. Celá problematika otevřených dat je též součástí vládní strategie pro boj proti korupci.



Obrázek 7. - Webová aplikace Monitor

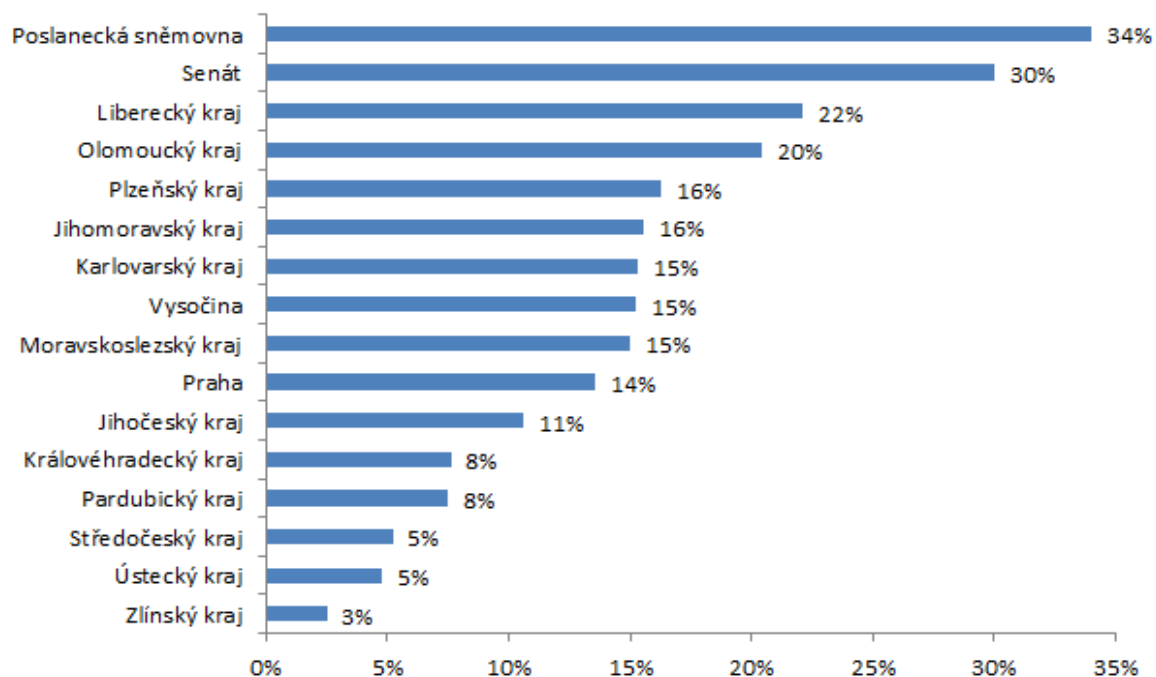
Zdroj: <http://monitor.statnipokladna.cz/2013/statni-rozpocet/kapitola/307>

2.4 Regionální úroveň

Na regionální úrovni veřejné správy se objevuje několik institucí, které se snaží s různou intenzitou podporovat myšlenku otevřených dat. Regionální úroveň, ať už v ČR či jinde ve světě je většinou v oblasti otevřených dat o něco napřed před státní úrovní. K tomu dochází především díky menšímu objemu informací, snadnější administrativě a snadnějšímu schvalovacímu procesu. Naopak problémem zde bývá větší nesourodost dat mezi jednotlivými subjekty.

2.4.1 Kraje

Na Obrázku 8 je k vidění srovnání otevřenosti zastupitelstev jednotlivých krajů, statistika je tvořena především ze zápisů z jednání, video či audio záznamů jednání, výsledků hlasování, usnesení ze schůzí a dalších. K hodnocení byl použit indikátor Open Data Index, vytvořený britskou organizací Open Knowledge Foundation.



Obrázek 8. - Srovnání otevřenosti zastupitelstev krajů

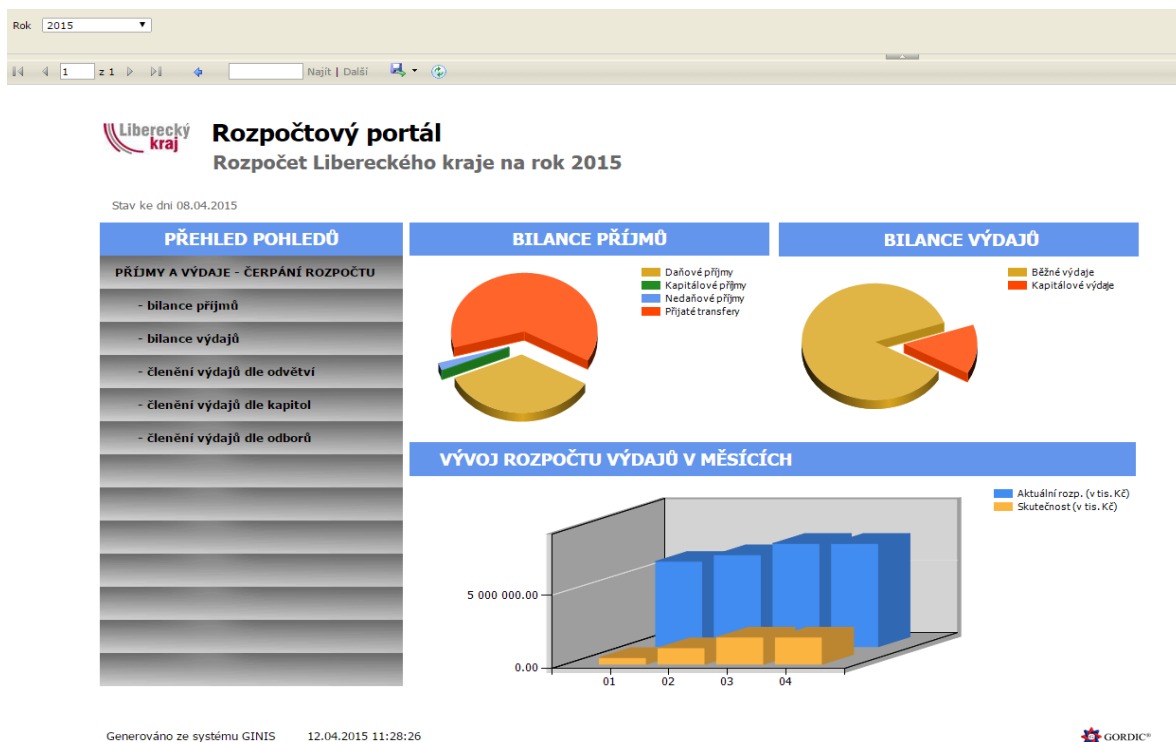
Zdroj: Gregor K., Datová otevřenost krajských zastupitelstev

Jak je ze srovnání patrné, krajská zastupitelstva jsou poměrně málo otevřená. V případě výborů by hodnoty otevřenosti byly ještě mnohem nižší. Jelikož z právní stránky téměř nic zastupitelstvům v otevřenosti nebrání je nutné slabé výsledky přičíst spíše technickým či jiným problémům.

2.4.2 Liberecký kraj

Za nejotevřenější kraj by bylo možné označit kraj Liberecký. Tento kraj získal v roce 2014 Cenu ministra vnitra za podporu otevřených dat. V době udělení ceny tvořily dokumenty zveřejněné Libereckým krajem na portále www.portal.gov.cz přibližně 40 % všech dat zveřejněných na tomto portále.

Liberecký kraj má na svých webových stránkách www.kraj-lbc.cz k dispozici aplikaci pro grafické znázornění rozpočtu s možností zobrazení jednotlivých položek, tzv. rozklikávací rozpočet:



Obrázek 9. - Rozklikávací rozpočet Libereckého kraje

Zdroj: <http://rozpocet.kraj-lbc.cz/>

2.4.3 Kraj Vysočina

Dalším krajem, kde se poměrně dobře daří prosazovat myšlenku otevřených dat je Kraj Vysočina. Přístup k otevřeným datům Kraje Vysočina je možný přes web <http://opendata.kr-vysocina.cz/>. Nachází se zde data o životním prostředí, veřejných zakázkách, kulturních událostech, různé geodatasety či městská telekomunikační síť. Datové zdroje jsou přidávány Odborem informatiky Krajského úřadu Kraje Vysočina. V lednu 2015 bylo na těchto stránkách k dispozici 16 datasetů.

Název Datasetu	Popis	Správce	ID Datasetu	Home
Akce v kraji Vysočina	Seznam akcí v kraji Vysočina - možnost výběru pomocí parametrů WWW služby. Popis exportu: http://extranet.kr-vysocina.cz/zlakovani.htm	Milan Lysa	vysocina_akce	
Aktuality ROWANet	RSS kanál s provozními aktualitami sítě ROWANet (novinky, výpadky)	Petr Tránek	vysocina_row	
Database usnesení zastupitelstva Kraje Vysočina	Databáze usnesení a okolů zastupitelstva Kraje Vysočina	Marie Šestáková	vysocina_zk	
Ekonomika příspěvkových organizací	Ekonomické výsledky jednotlivých příspěvkových organizací. Možnost výběru období a skupin organizací	Michal Narbonne	vysocina_ekonomika_po	
Metropolitní síť	Optická telekomunikační síť Kraje Vysočina představuje soubor optických sítí, které umožňují vedení datových toků prostřednictvím optických vláken. Optické sítě, kterými jsou vedeny data na území kraje jsou tyto: Rowanet, Dial Telecom, GTS Novera, JME optika, Tranogas, Metropolitní síť KrÚ Jihlava, Net4Net, Opticon, Optikon kos, Pragonet, První telefonní, Selfnet, O2, Teta Sonera, Tiscali optika, ČD Metabika, ČD Telematika a Pegas burley.	ad@telecom.cz	vysocina_metropolitni	
Památky	Datová vřstiva zobrazuje zákonem protážená plošně chráněná památková území, která vykazují kulturní hodnoty. Jedná se o památkové rezervace (městské, vesnické), památkové zóny (městské, vesnické, krajinné) a ochranná pásma.	ad@telecom.cz	vysocina_pamatky	
Příspěvkové organizace	Seznam příspěvkových organizací Kraje Vysočina	Matěj Čep	vysocina_prispevkove_organizace	
RSS služby krajského noviny	RSS kanály agregují novinky z různých částí webu www.kr-vysocina.cz .	Webmaster KÚ	vysocina_web	

Obrázek 10. - Portál otevřených dat kraje Vysočina

Zdroj: <http://opendata.kr-vysocina.cz/>

2.4.4 Města

Popsat situaci otevřených dat v českých městech je poměrně složitá záležitost. Většina měst nemá pro otevřená data vytvořen speciální web. Většina měst, která zveřejňuje otevřená data, je zveřejňuje nahodile a občas možná i bez vědomí, že se o otevřená data jedná. Přesto se některá města snaží tvořit na svých webech ucelené sekce s otevřenými daty. Mezi prvními městy, která otevřela svá data, jsou například Děčín či Opava. V dubnu 2015 byla také na portálu <http://www.geoportalpraha.cz/cs/opendata> zveřejněna geografická data hlavního města Prahy.

Dle nominací v soutěži Otevřeno x Zavřeno, pořádané Otevřenou společností v roce 2014 se následující města různou měrou podílela na otevřených datech: Děčín, Opava, Zlín, Liberec, Uherský Brod, Hradec Králové, Mníšek pod Brdy, Kolín a další. [12]

Města, která naopak učinila negativní kroky v oblasti otevřených dat: Pardubice (lhaní o plánech staveb), Poděbrady (rada města odepírá poskytnout zápisy ze svých jednání, přijímá usnesení, k nimž nemá kompetenci a ignoruje rozhodnutí krajského úřadu i stanoviska Ministerstva vnitra), Lanškroun (obstrukce při poskytování informací o soudních sporech) a další. [12]

2.4.5 Město Děčín

Město Děčín je možné označit za jakéhosi průkopníka v oblasti otevřených dat mezi českými obcemi. Město díky spolupráci s Fondem Otakara Motejla a Fórem pro otevřená data svá data zveřejnilo mezi prvními v České republice. S Fondem Otakara Motejla město též podepsalo Memorandum o spolupráci při prosazování otevřených dat. V soutěži Otevřeno x Zavřeno, pořádané Otevřenou společností, získalo město Děčín v roce 2014 zvláštní cenu za otevřená data.

Otevřená data města Děčín jsou pro uživatele dostupná na webu http://mmdecin.cz/dokumenty/cat_view/238-otevrena-data. Hned v úvodu je pro uživatele k dispozici informace o tom jak je možné zveřejněná data využívat, kopírovat či dále šířit:

Data zveřejněná na webových stránkách statutárního města Děčín a označená jako otevřená data smí uživatel:

- šířit – kopírovat, distribuovat a sdělovat veřejnosti,
- využít a citovat,
- využívat komerčně a to při dodržení podmínky, že jako zdroj dat musí být uvedeno statutární město Děčín.

Uživatel dále bere na vědomí, že:

- v případě dalšího šíření otevřených dat statutárního města Děčín je povinen uvést znění těchto podmínek užití, nejlépe přímým odkazem na tuto webovou stránku,
- tyto podmínky užití neplatí pro dokumenty, na něž se vztahují práva duševního vlastnictví třetích stran,
- statutární město Děčín se zříká odpovědnosti za případné škody vzniklé dalším užitím a zpracováním otevřených dat, jakož i odpovědnosti za škody způsobené případnými chybami, které by otevřená data mohla obsahovat. [13]

Stránka Otevřená data je rozdělena do několika kategorií a podkategorií. Samotná data jsou zde zveřejněna v různých formátech, např.: *.ods, *.xlsx, *.csv, *.pdf, *.jpg a další. Jedná se o data z rozpočtů, majetku města, účetnictví či geodata. Na webu jsou též k dispozici

odkazy na kompletní videozáznamy (nahrané na webu www.youtube.com) a zápisy ze zasedání zastupitelstva města.

2.5 Legislativa v ČR

V každé zemi panuje rozdílný přístup k otevřeným datům. Pro ČR, stejně jako pro ostatní členské země EU, jsou klíčová nařízení a směrnice Evropské Komise, které musí zapracovat do svých národních zákonů. Samotný přístup k informacím je pro občany zaručen listinou základních práv a svobod.

2.5.1 Zveřejňovaná data

Zákon o státní službě

V rámci strategie „Akční plán České republiky Partnerství pro otevřené vládnutí na období let 2014 až 2016“ byl v listopadu 2014 schválen zákon o státní službě. Plné účinnosti zákon nabyl 1. ledna 2015. Mezi klíčové body tohoto zákona patří odpolitizování státní správy, jmenování náměstka pro státní službu a jmenování státních tajemníků v ministerstvech a v Úřadu vlády České republiky. 28. ledna 2015 byl náměstkem pro státní službu jmenován Josef Postránecký.

Zákon číslo 106/1999 Sb., o svobodném přístupu k informacím

Zákon byl vytvořen na základě předpisu Evropského společenství. Později byl doplněn několika novelami na základě nově vydaných směrnic Evropského parlamentu.

Citováno přímo ze zákona:

„Povinnými subjekty, které mají podle tohoto zákona povinnost poskytovat informace vztahující se k jejich působnosti, jsou státní orgány, územní samosprávné celky a jejich orgány a veřejné instituce.“ [14]

Úřady mohou informace zveřejňovat buď aktivně samy, například na internetu, nebo na vyžádání.

„Žadatelem pro účel tohoto zákona je každá fyzická i právnická osoba, která žádá o informaci.“ [14]

V první polovině roku 2015 se v parlamentu ČR nachází novela tohoto zákona přijímající změny na základě směrnice Evropského parlamentu a Rady č. 2013//37/EU ze dne 26. června 2013. Tato novela do agendy zpřístupňování informací zapracovává myšlenku otevřených dat, přesněji řečeno nabádá k zveřejňování dat ve strojově čitelném formátu. Jedná se tedy spíše o technické změny, které by ovšem v konečném důsledku měly vést k efektivnějšímu využití zveřejňovaných informací.

Zveřejňování záznamů z jednání zastupitelstev

Obce nemají ze zákona žádnou povinnost zveřejňovat záznamy z jednání zastupitelstev. Obce, které se na základě zákona číslo 106/1999 Sb., o svobodném přístupu k informacím, rozhodnou dobrovolně zveřejňovat svá data, tak musí dbát především na odstranění osobních údajů, výjimkou je situace, kdy dotčená osoba předem souhlasí se zveřejněním svého jména. Též je možné zveřejnit osobní údaje, které je možné zveřejnit na základě jiných zákonů a často již někde zveřejněny jsou, například v obchodním rejstříku.

2.5.2 Nezveřejňovaná data

Zákon číslo 101/2000 Sb., o ochraně osobních údajů a o změně některých zákonů

Dle zákona o ochraně osobních údajů, jak již bylo výše zmíněno, není možné až na výjimky zveřejňovat data osobního charakteru. V této souvislosti může dojít i k tzv. problému nechtěné identifikace, kdy se z dat odstraní např. osobní údaje, ale pomocí kombinace ostatních dat z datasetu s jinými daty lze danou osobu identifikovat. Při zveřejňování dat je třeba na toto brát zřetel a pečlivě zvážit, co zveřejnit.

Stejně není z principu možné zveřejňovat strategická data, např. data o rozvodech elektrické sítě, strategická data ČR či jiná ze své podstaty tajná či důvěrná data.

Další kategorií, na kterou je třeba brát ohled, jsou autorská díla. Tuto oblast řeší zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů.

2.6 Zahraničí

Otevřená data se v zahraničí prosazují hlavně ve vyspělých zemích. V rámci boje proti korupci se ovšem otevřená data mohou prosadit i v některých rozvojových zemích. Poměrně překvapivě může znít i informace, že určitá snaha o otevřená data je k vidění i v jinak velmi cenzurované Číně či například v Keni.

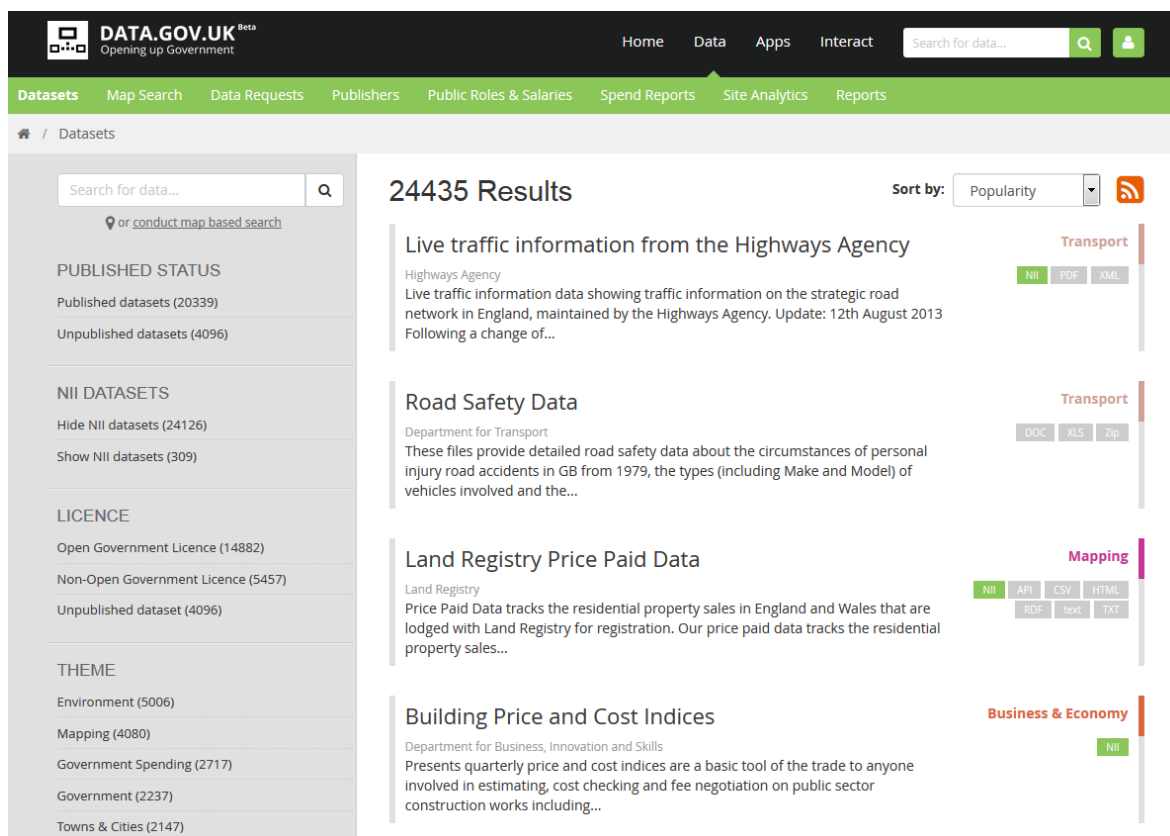
2.6.1 Evropa

V Evropě je možné označit všechny země za relativně otevřené. Nejhůře jsou na tom v tomto ohledu pravděpodobně Bělorusko a Ukrajina. Naopak za nejvíce otevřenou zemi je možné označit Velkou Británii, která vládne i většině světových statistik otevřenosti spolu se Spojenými Státy.

Dle studie MGI by otevřená data mohla Evropskému veřejnému prostoru ročně přinést až 250 miliard Euro. [15]

Velká Británie

Velká Británie je v průběhu několika posledních let považována za vzor v oblasti problematiky otevřených dat. Britský datový katalog v současné době obsahuje přibližně 8 500 datasetů. Veškerá zveřejněná vládní data jsou pro britské občany dostupná na portálu data.gov.uk. Jsou zde k dispozici všechna data z hlavních vládních sektorů ale i některých lokálních úřadů a institucí. Některá z dat zde shromážděných již byla dříve dostupná na jiných místech, nicméně zde jsou všem dostupná v ucelené formě s lepší možností vyhledávání.

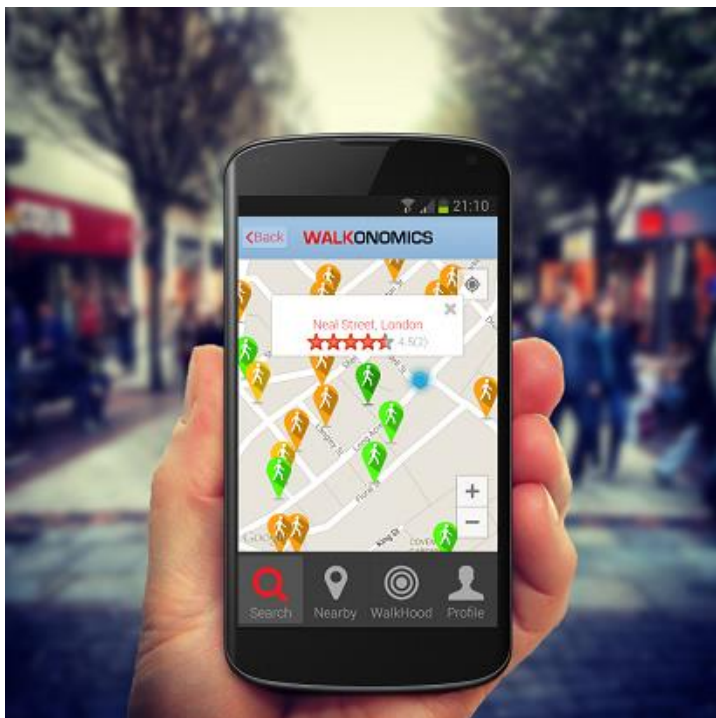


Obrázek 11. Britský portál pro otevřená data - data.gov.uk

Zdroj: <http://data.gov.uk/data/search>

Hlavním přínosem portálu by dle jeho zakladatelů mělo být využití surových dat k následnému vytvoření uživatelsky příjemnějších aplikací v jednotlivých oblastech. Na webu jsou tak k dispozici nejenom samotná data, ale i aplikace, která tato data využívají. Tato forma je pro většinu běžných uživatelů příjemnější než hledání v tisících řádcích tabulek. V únoru 2015 bylo na tomto portále k dispozici 367 aplikací využívající otevřená data.

Například aplikace Walkonomics pomocí souboru několika údajů (množství chodců, kvalita povrchu, šířka chodníku, místní kriminalita atd.) zobrazuje jak snadná je chůze v různých ulicích. Při pohledu na mapu si tak uživatel může vytvořit ucelenější pohled než při prohlížení jednotlivých tabulek.



Obrázek 12. - Walkonomics - aplikace pro zobrazení průchodnosti ulic

Zdroj: http://data.gov.uk/sites/default/files/files/apps/Walkonomics_v2_Nexus_in_Street.png

Britská vláda má pro projekt data.gov.uk vytvořený speciální tým, který dbá na to, aby data byla napříč všemi resorty dodávána včas a ve správné formě. Zároveň na tomto projektu pracuje mnoho předních technických odborníků.

Rakousko

Jednou ze zemí, od kterých se snaží Česká republika v oblasti otevřených dat inspirovat, je Rakousko. Stejně jako v ČR se otevřenými daty v Rakousku začaly nejdříve zabývat lokální oblasti, především města. V červenci 2011 založilo Federální Kancléřství společně s městy Vídeň, Linz a Graz iniciativu Otevřená data Rakousko. Za účasti mezinárodních expertů byly dohodnuty společné standardy a podmínky pro rozvoj a přínos pro všechny zúčastněné strany. Většina z těchto principů byla přejata od amerických pravidel pro otevřená vládní data. Federální Kancléřství poskytlo k projektu potřebnou IT infrastrukturu. Pro jednotlivá města je v Rakousku zaveden jednotný systém domén, např.: data.salzburg.gv.at, data.linz.gv.at, data.graz.gv.at, atd.

Podobně jako Velká Británie má Rakousko svůj centrální web pro vládní data – data.gv.at.

Aktuell: ViennaGIS® verschenkt Geodaten-können wir uns das leisten? [API](#)

Suchbegriff (z.B. Finanzen, Wahlen) [Suche starten](#)

[Datenkatalog](#) [Apps & News](#) [→ Katalog durchstöbern](#)

data.gv.at – offene Daten Österreichs

[Startseite](#) [Katalog](#) [Anwendungen](#) [Linked Data](#) [News](#) [Infos](#) [Netiquette](#) [Kontakt](#)

offene Daten Österreichs – lesbar für Mensch und Maschine

Vielfalt, Transparenz, Offenheit, Demokratie

data.gv.at bietet einen [Katalog offener Datensätze und Dienste](#) aus der öffentlichen Verwaltung, welche auf den [Open Data-Prinzipien](#) basieren.

Sie können diese Daten frei nutzen – zur persönlichen Information und auch für kommerzielle Zwecke wie [Applikationen oder Visualisierungen](#). Details hierzu finden Sie im Menüpunkt [Netiquette](#).

Mehr Hintergrundinfos erhalten Sie auch im [Video "Was ist Open Data?"](#)

Welche Daten sind verfügbar?
Die österreichische Verwaltung bietet alles von Wetterdaten bis zu genauen statistischen Daten aus allen Ressorts und Verwaltungsebenen.
[→ Katalog](#)

Was kann man damit machen?
Die Daten stehen im Rahmen der Open Government Data frei zur Verfügung. Genauere Datenrichtlinien erfahren Sie bei den einzelnen Datensätzen.
[→ Anwendungen](#)

Was tut sich im Bereich OGD?
Aktuelles zu Innovationen, Veranstaltungen, Zusammenarbeit, Technologie, Anwendungen im Bereich der Offenen Daten in Österreich.
[→ News](#)

Wer steckt dahinter?
Die Cooperation OGD Österreich aus Bund, Ländern und Gemeinden, die mit data.gv.at eine Plattform für offene Daten aus der Verwaltung anbietet.
[→ Cooperation OGD Österreich](#)

Obrázek 13. - Rakouský portál pro otevřená data - data.gv.at

Zdroj: <https://www.data.gv.at/>

Estonsko

Estonsko se v 90. letech po rozpadu Sovětského svazu nacházelo ve složité situaci a hrozila mu vleklá hospodářská krize. V této době se Estonci rozhodli svojí plánovanou ekonomiku přeměnit na vzdělanostní s důrazem na nové technologie a vysokou přidanou hodnotu. Byla vystavěna síť informačních služeb, která umožňuje efektivní komunikaci mezi úřady a občany. Estonsko není a nikdy nebude průmyslový obr, přesto je díky propracovanosti služeb souhrnně označovaných jako e-government vzorem pro mnoho zemí v regionu.

Evropská Unie

Evropská Unie se především pomocí svých předpisů a nařízení snaží jednotlivé země přimět k přijetí zákonů, které by dávaly za povinnost státním orgánům zveřejňovat data veřejného zájmu a občanům k těmto datům umožnily a usnadnily přístup.

Hlavním dokumentem je směrnice 2003/98/ES o opakovaném použití informací veřejného sektoru a novelizovaná směrnice 2013/37/EU ze dne 26. června 2013. Směrnice stanoví, že členské státy do 18. července 2015 přijmou a zveřejní právní a správní předpisy nezbytné pro dosažení souladu s touto směrnicí. Hlavními body novelizované směrnice jsou:

- **Obecná zásada**

Data, na které se směrnice vztahuje, musí být v souladu s podmínkami směrnice opakovaně použitelná pro komerční i nekomerční účely. [16]

- **Strojová čitelnost**

„Subjekty veřejného sektoru poskytují své dokumenty v jakémkoliv již existujícím formátu nebo jazyce, a pokud je to možné a vhodné, v otevřeném a strojově čitelném formátu spolu s jejich metadaty. Formát i metadata by měla co nejvíce splňovat formální otevřené normy“. [16]

2.6.2 Svět

V minulých letech se mnoho zemí začalo zabývat myšlenkou otevřených dat. Zvláště pro rozvíjející se země jsou otevřená data jednou z relativně jednoduchých možností jak zlepšit státní správu a snížit obvykle velké množství korupce. Díky lepšímu kapitálu v oblasti informačních technologií a tradičně silnější demokracii se stejně jako v Evropě i ve světě otevřená data prosazují především ve vyspělých zemích.

„Podle studie McKinsey mohou otevřená data ušetřit vyspělým západním ekonomikám 100 miliard euro ročně na straně institucí a dalších 450 miliard u jejich klientů.“ [17 s. 12]

V případě této studie je však nutné podotknout, jak je uvedeno v ní samotné, že se jedná pouze o odhad a k přesnějším údajům chybí paradoxně právě více dat.

USA

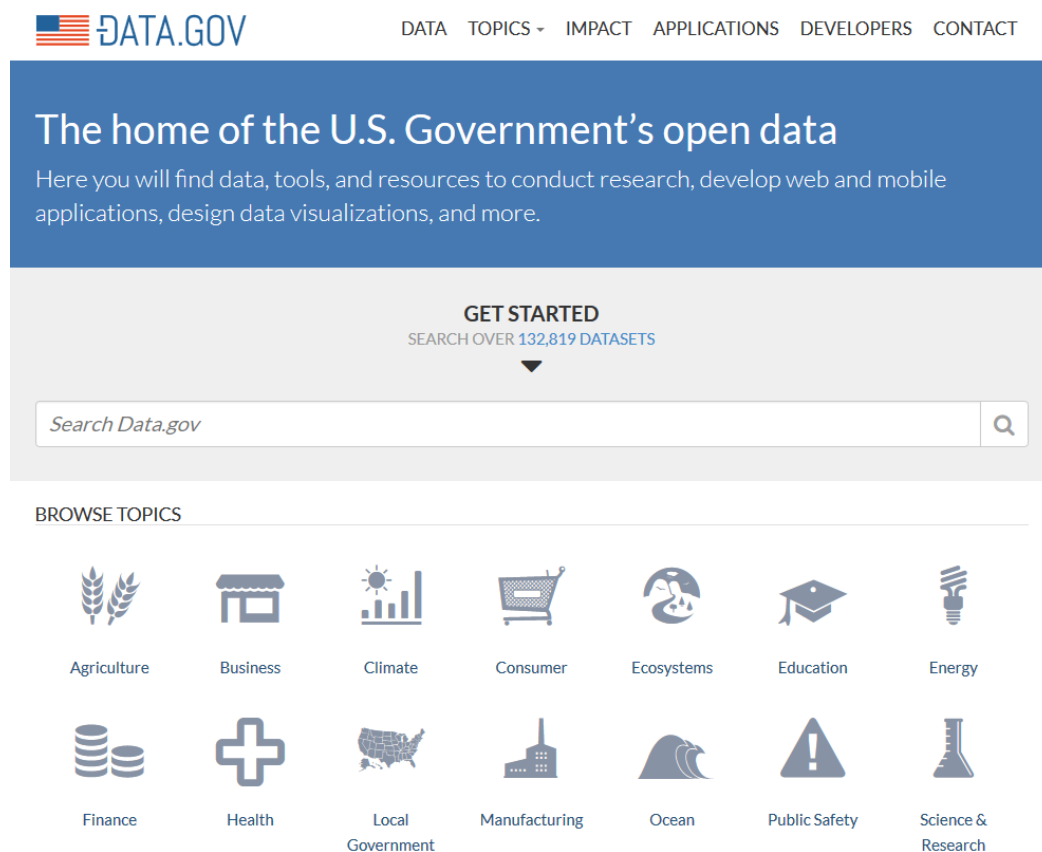
Jako první se problematikou otevřených dat začala zabývat vláda USA. Díky podpoře prezidenta Obamy, který nastoupil do funkce v roce 2009, získal projekt otevřených dat rychle na významu a postupně se začali přidávat další země. Hlavní myšlenkou bylo posílit demokracii a zlepšit fungování státu.

21. ledna 2009 řekl prezident Obama ve svém projevu následující:

„Moje administrativa se zavazuje k vytvoření bezprecedentní úrovně otevřenosti ve vládě. Budeme společně pracovat na zajištění důvěry veřejnosti a vytvoření systému transparentnosti, zapojení veřejnosti a spolupráce. Otevřenost posílí naši demokracii a zvýší efektivnost a účinnost vlády. Vláda by měla být transparentní. Transparentnost zvyšuje spolehlivost a poskytuje občanům informace o tom, co jejich vláda dělá.“

[18 s. 42]

Podobně jako Velká Británie má vláda USA k dispozici centrální uložisko otevřených dat dostupné přes web data.gov.



Obrázek 14. - Americký portál pro otevřená data - data.gov

Zdroj: <http://www.data.gov/>

Čína

Čína je známá jako velmi cenzurovaná země, přesto i zde jsou k nalezení jedinci, kteří se toto snaží změnit. V roce 2013 Čína přistoupila prostřednictvím svých ambasadorů k projektu OKFN, nicméně od té doby nebylo na vládní úrovni učiněno mnoho kroků k vylepšení situace otevřených dat v Číně. Jednou ze světých výjimek je čínský Národní statistický úřad. Místem, kde se v Číně prosazují otevřená data nejvíce, jsou tak dnes především jednotlivá města. Asi nejpopulárnějším využitím otevřených dat v Číně jsou smogové mapy s aktuální smogovou situací. Je ovšem nutné dodat, že vzhledem k tomu, že veškerá data v Číně mohou být kontrolována vládou, není možné těmto datům dávat plnou důvěru.

2.7 Organizace a projekty v zahraničí

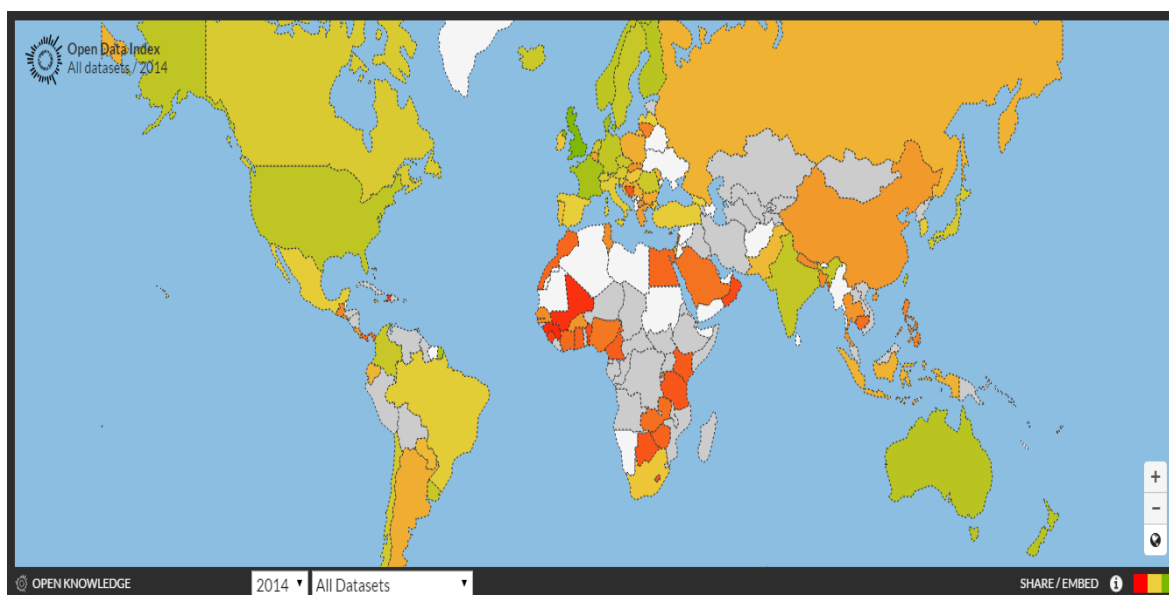
Existuje mnoho zahraničních i mezinárodních organizací zabývajících se přímo i nepřímo problematikou otevřených dat.

OKFN - Open Knowledge Foundation

Mezinárodní nezisková organizace odborníků zabývajících se sdílením znalostí a otevřenými daty. Sídli ve Velké Británii. Principy fungování OKFN:

- mezinárodní síť lidí prosazujících otevřenost,
- monitorování otevřenosti ve světě,
- poskytování informací a vzdělávání lidí v oblasti otevřenosti,
- správcovství a pomoc s projekty,
- služby a konzultace. [19]

Organizace OKFN vydává každý rok celosvětový index otevřenosti – Global open data index. V jednotlivých zemích jsou sledovány faktory otevřenosti v následujících oblastech: výsledky voleb, národní mapy, registry firem, poštovní směrovací čísla, národní statistiky, vládní výdaje, vládní rozpočet, veřejná doprava, legislativa, znečištění.

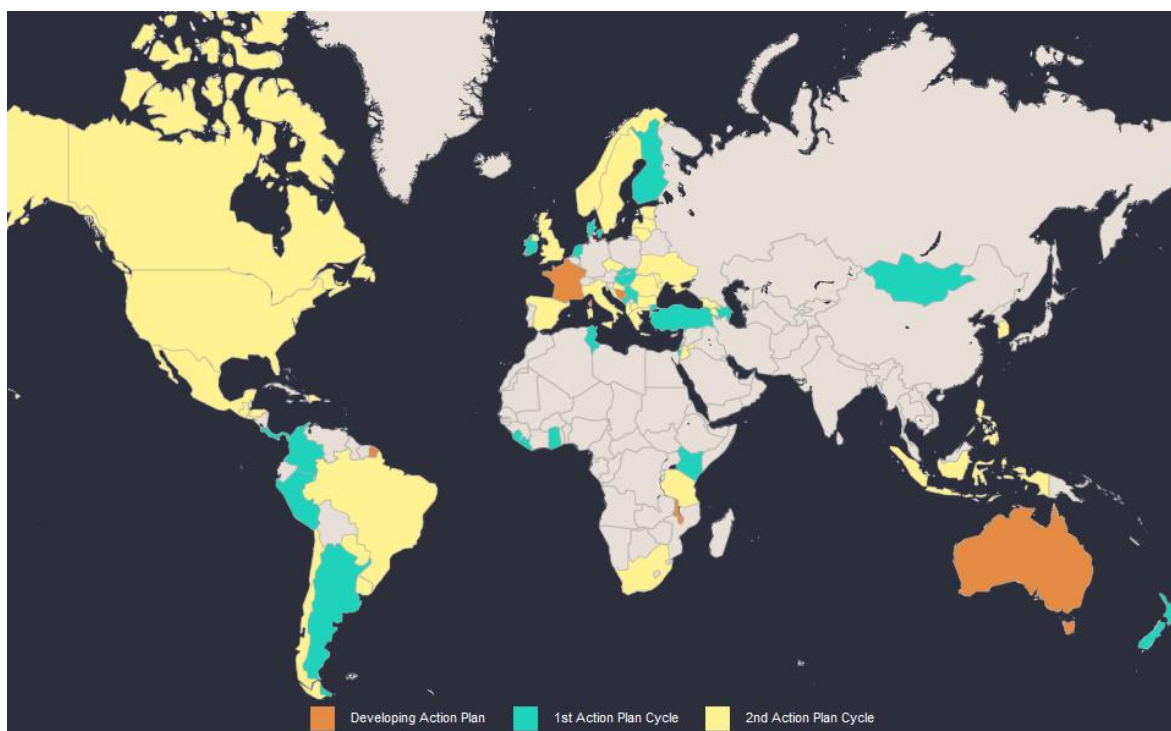


Obrázek 15. - Global open data index

Zdroj: <http://index.okfn.org/>

Mezinárodní iniciativa administrativy USA podporující otevřenost, transparentnost a boj proti korupci. Jedná se o projekt, kde se jednotlivé vlády participujících zemí zavazují ke krokům vedoucím ke zlepšení v jednotlivých oblastech. Vlády mají poté povinnost iniciativu informovat o tom, jak se jí daří dané závazky plnit. K této iniciativě přistoupila i ČR usnesením vlády ze dne 14. září 2011.

Na Obrázku 16 je možné vidět země, které se rozhodly připojit k mezinárodní iniciativě OGP.



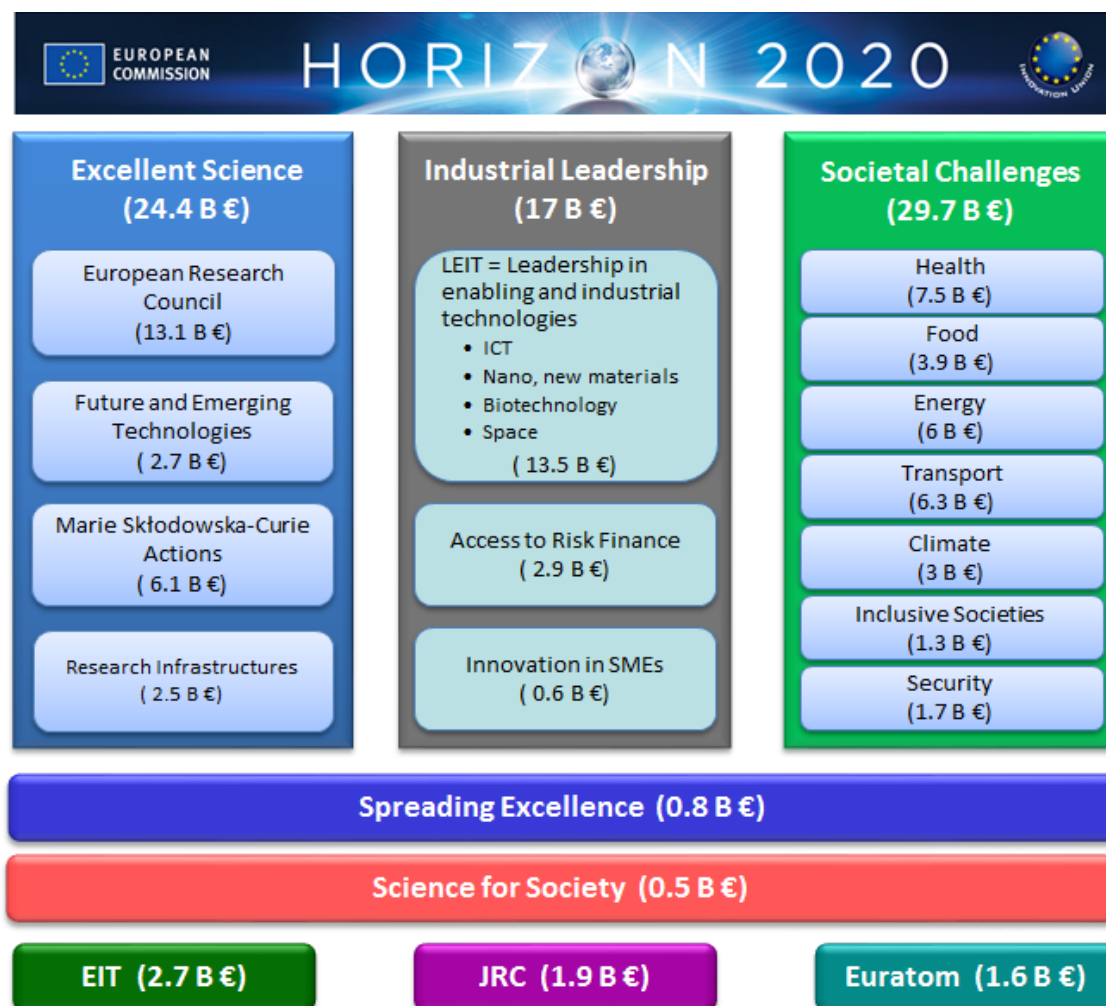
Zdroj: www.opengovpartnership.org/countries

Fórum, jež má za cíl podpořit občanské iniciativy řešící problematiku otevřených veřejných institucí. Na svém webu www.openingparliament.org má vystavenou „Deklaraci parlamentní otevřenosti“, kterou v současné době podporuje více než 140 organizací ze 75 zemí světa. Zároveň jí přímo podporuje i několik národních parlamentů. Deklarace ve 44 člancích uvádí, co vše by měly parlamenty splňovat, aby mohly být označeny za otevřené.

Např. Článek 19: „Záznamy z jednání parlamentních výborů včetně dokumentů vytvořených a přijatých, výpovědi svědků veřejných schůzí, stenozáznamy a záznamy konkrétní činnosti ve výborech by měly být vždy veřejně publikovány.“

Horizon 2020

Jedná se o rámcový program na evropské úrovni pro financování vědy, výzkumu a inovací pro roky 2014-2020. Historicky navazuje na již 7. rámcový program pro výzkum, technologický rozvoj a inovace (2007-2013). Nový program klade větší důraz na inovace, financování úvěrů u malých a středních podniků, snížení administrativní zátěže a snahu o lepší převedení výsledků výzkumů do praxe. Celkově by měl tento program v průběhu 7 let rozdělit téměř 80 miliard Euro.



Obrázek 17. - Rozložení rozpočtu Horizon 2020

Zdroj: <http://cerneu.web.cern.ch/sites/cerneu.web.cern.ch/files/Horizon%202020%20pillars.png>

Na rozdíl od 7. rámcového programu s sebou Horizon 2020 přináší povinnost poskytovat otevřený přístup k vědeckým publikacím, které jsou považovány za výstup výzkumu/projektu, ve všech oblastech. Kromě samotných publikací musí do repozitářů uložit též metadata k publikaci, která následně využívají jiné nástroje a umožňují tak efektivní vyhledávání a vyhodnocování. Pravidla pro zveřejňování výsledků výzkumů se řídí nařízením EP a Rady č. 1290/2013 ze dne 11. prosince 2013. Stejně tak toto nařízení řeší práva duševního vlastnictví a licence. Jsou zde stanoveny 2 základní cesty k otevřenému přístupu:

1. **Zelená cesta** – článek je do repozitáře uložen přímo autorem, tedy klasická cesta.
2. **Zlatá cesta** – článek zveřejňuje vydavatel, obvykle žádá po autorovi poplatek za zpracování (neplatí pro univerzitní, vědecké či neziskové organizace), v rámci programu Horizon 2020 je možné si tyto náklady přidat do uznatelných nákladů.

Pro obě cesty platí, že povinnost zajištění otevřeného přístupu musí být splněna do 6 měsíců od publikování výsledků (respektive 12 měsíců v případě výzkumů z humanitní oblasti).

Existují 3 základní typy repozitářů, které mohou instituce využít:

- repozitář vlastní organizace,
- repozitář oboru dané výzkumné oblasti,
- centrální repozitáře – například repozitář ZENODO (repozitář vzniklý v rámci projektu OpenAIRE), který byl vybudován, aby usnadnil autorům povinnosti v rámci Horizon 2020.

Pro vybrané oblasti výzkumu je v programu Horizon 2020 zařazen pilotní projekt, ve kterém musí výzkumníci poskytnout otevřený přístup i k výzkumným datům. Dobrovolně se k tomuto projektu mohou připojit i projekty z jiných oblastí. Kromě toho Horizon 2020 doporučuje zveřejňovat i ostatní data, týkající se daného výzkumu, např. záznamy z konferencí, prezentace atd.

Share-PSI 2.0

Evropský projekt zaměřený na výměnu nápadů a zkušeností týkajících se otevřených dat. Tento projekt je tvořen partnery ze 46 různých oblastí zahrnující instituce veřejného sektoru, školy a univerzity a další komerční či nekomerční organizace. Financování projektu je z části zajištěno Evropskou komisí v rámci programu ICT Policy Support (ICT PSP), který je součástí Rámcového programu pro konkurenceschopnost a inovace (CIP).

LAPSI 2.0

Cílem projektu bylo identifikovat zbývající právní bariéry a překážky v oblasti přístupu a využívání informací veřejného sektoru v rámci Evropského prostoru a zároveň navrhnout metriky a nástroje ke stimulaci pokroku Evropského trhu směrem k otevřeným datům. Na projektu se podíleli experti z veřejného i soukromého sektoru 15 zemí. Byl financován Evropskou komisí v rámci Rámcového programu pro konkurenceschopnost a inovace (CIP). Projekt byl ukončen 28. listopadu 2014 svou závěrečnou konferencí v Bruselu, kde byla prezentována navrhovaná doporučení. [20]

ePSIplatform

Iniciativa Evropské Komise pro podporu dynamických informací veřejného sektoru (PSI) a znovuvyužití otevřených dat napříč Evropskou unií. Na portále epsiplatform.eu jsou k dispozici zprávy, reporty a případové studie ze všech zemí Unie. Portál umožňuje expertům z jednotlivých zemí sdílet své znalosti a zkušenosti s implementací otevřených dat v jejich národních prostředích.

2.8 Soukromý sektor

Soukromý sektor může stejně jako veřejná správa zveřejňovat svá data. Většinou však nastává situace, kdy naopak soukromý sektor čerpá z dat, která byla zveřejněna jako „otevřená“ veřejnou správou. Díky tomu mohou otevřená data přinést benefit jak veřejné správě (např. úspora finančních prostředků díky veřejným zakázkám), tak i soukromým subjektům, které by jinak mohly být znevýhodněny oproti ostatním subjektům (např. korupčním jednáním).

Jedním ze známých využití otevřených dat v soukromém sektoru je například zpřístupnění amerického armádního systému GPS. Tento systém původně sloužil pouze vojenským účelům a pro veřejnost se otevřel až během devadesátých let. Od té doby se rozrostl do obrovského byznysu s miliardami dolarů. Dalším příkladem může být například zpřístupnění meteorologických dat, kdy různé firmy na těchto datech staví své komerční předpovědní modely.

Harvey Lewis, analytik společnosti Deloitte UK rozděluje komerční možnosti využití otevřených dat na následující obchodní modely:

1. Suppliers – dodavatelé

Suppliers neboli dodavatelé jsou ti, kteří samotná otevřená data poskytují a nemusí to ani být primárním cílem jejich činnosti nebo předmětem přímého zisku.

Např. telekomunikační operátor, který poskytuje datové zdroje o svých cenových tarifech pro nezávislé srovnávače, čímž si buduje vůči zákazníkům lepší pozici – odlišuje se tím od konkurence, která cenové tarify umně skrývá v pralese svých nepřehledných ceníků.

2. Aggregators – sběrači

Tyto podniky data sbírají a zpracovávají. Přidanou hodnotu vytváří jejich zpracováním, čištěním nebo kombinováním.

Typickým příkladem agregátora dat je projekt OpenCorporates.com. Jedná se o největší světovou databanku údajů o firmách, která do své databáze stahuje data nejen z několika desítek obchodních rejstříků světa, ale také z dalších dostupných zdrojů.

3. Developers – vývojáři

Pravděpodobně nejčastější model, jedná se o aplikace mobilní, webové nebo třeba poskytované prostřednictvím kiosků či informačních terminálů, jejich cílovou skupinou jsou nejčastěji koncoví zákazníci.

4. Enrichers – obohátelé

Společnosti, které k datům přidávají i svou vlastní expertízu a teprve tu nabízí jako službu klientům. Jedním z nejzářivějších příkladů je startup ze San Diega s názvem BrightScope. Věnuje se porovnávání poplatků za správu penzijních plánů amerických firem (tzv. 401k plans).

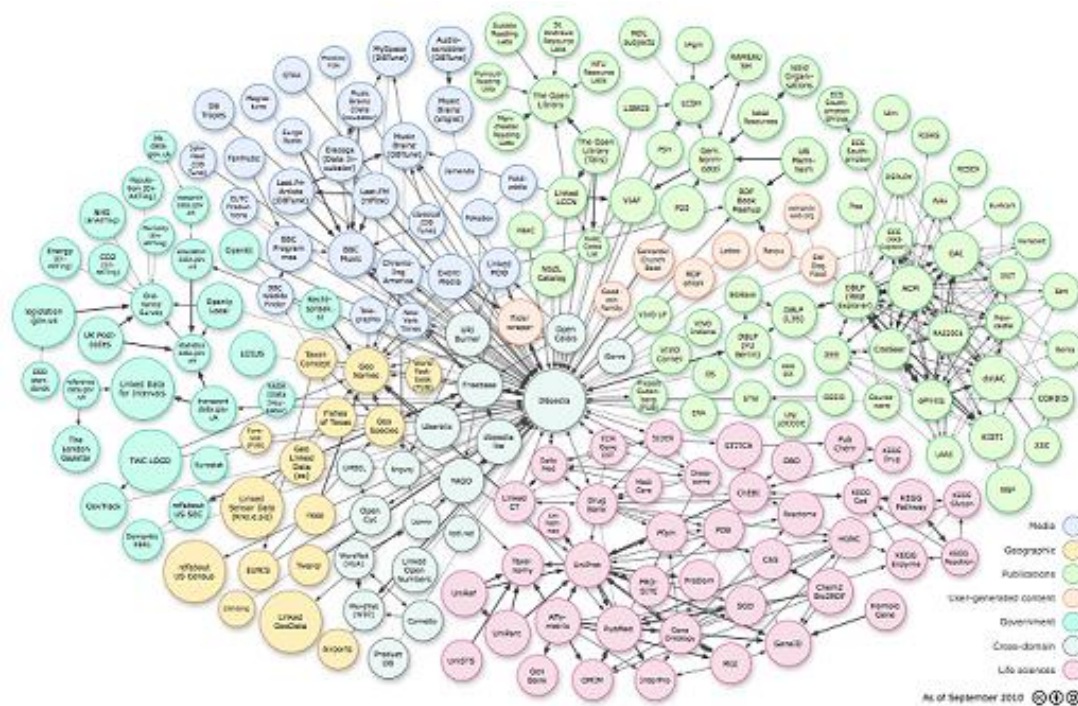
5. Enablers – umožňovatelé

Jejich cílem je přinášet nástroje, metody či technologie, které otevřená data umožní zpřístupňovat a zpracovávat.

Jedním z největších hráčů na tomto poli je americká společnost Socrata, která poskytuje vládám či municipalitám softwarová řešení i potřebné procesní i právní konzultace po zavádění otevřených dat. Její platformu využívá například město Chicago, stát Oregon nebo východoafrická Keňa. Do této kategorie se dají zařadit také nástroje pro práci s daty od firmy Google, jako jsou Google Fusion Tables, Google Refine nebo Google Public Data Explorer. [21], [22]

3 Propojitelná otevřená data

Aby data mohla být označena jako propojitelná otevřená data, musí splňovat přísnější podmínky než „obyčejná“ otevřená data. Jedná se v podstatě o čtyřhvězdičkovou nebo pětihvězdičkovou úroveň otevřenosti dat. Jak vypadá mapa takového webu otevřených propojitelných dat, ukazuje následující obrázek.



Obrázek 18. - Mapa propojení LOD

Zdroj: <http://lod-cloud.net/>

3.1 Definice LOGD

Stejně jako u klasických otevřených dat jde hlavně o data veřejné správy a vzniká tak tedy LOGD:

“LOGD are all stored data of the public sector connected by the World Wide Web which could be made accessible in a public interest without any restrictions for usage and distribution” [2 s. 268]

Neboli: propojitelná otevřená data veřejné správy jsou data veřejného sektoru propojená pomocí www, která by měla být dostupná veřejnému zájmu bez jakýchkoli omezení užití či šíření.

3.2 Výhody LOD

Oproti klasickým otevřeným datům přinášejí propojitelná otevřená data další výhody:

- dávají informacím lepší význam,
- nad daty je možné vytvářet dotazy,
- je možné vytvářet spojení s daty 3. stran,
- standardizované protokoly,
- věci jsou pojmenované pomocí URI.

V souvislosti s propojitelnými daty se vyskytuje pojem inference znalostí. Inference znamená vytváření znalostí/poznatků na základě propojení již známých znalostí/poznatků. Takové vytváření nebo hledání znalostí je v tomto případě možné jak manuálně tak strojově. [23]

V případě stroje se pak jedná o tzv. inference engine neboli počítač či algoritmus, který má schopnost logického odvozování. Stroj porovnává vstupní fakta se známými fakty a pravidly v dané oblasti expertízy a z toho vyvozuje závěry. Správně fungující inferenční stroj řeší problém podobným způsobem, jaký by použili specialisté v daném oboru. [24]

3.3 Aplikace a nástroje pro práci

LOD2 Stack

SW nástroj a zároveň hlavní cíl vzniklý v rámci projektu LOD2. Jde o SW, který integruje několik vzájemně propojených nástrojů, které dohromady pokrývají celý proces/životní cyklus dat – viz Obrázek 19.

Hlavní funkce LOD2 Stack:

- extrakce RDF z textu, XML nebo SQL,
- dotazování a výzkum za pomoci SPARQL,
- určování vlastnictví využitím sémantické Wiki,
- polo-automatické propojení mezi zdroji propojitelných dat,
- obohacení pomocí znalostní báze a opravy. [25]

3.4 LOD v ČR

Zpočátku je nutné říci, že v České republice v současné době většina otevřených dat nespĺňuje podmínky, aby mohla být označena též za data propojitelná. Nicméně je ČR zapojena do několika mezinárodních projektů pro otevřená propojitelná data, jako např. LOD2, Share-PSI 2.0 a COMSODE, blíže představené v kap. 2.7 a 3.5.

linked.opendata.cz

Jedním z míst zaměřených na otevřená propojitelná data je např. web linked.opendata.cz jako subdoména opendata.cz. V dubnu 2015 se zde nacházelo 13 datasetů, např. kontroly ČOI, registry smluv, registry územní identifikace, počty obyvatel nebo soubory identifikačních čísel.

Otevřená propojitelná data v oblasti veřejných rozpočtů

Jedná se o projekt vedený Dušanem Chlapkem v období od ledna 2014 do září 2015. Jedním z hlavních úkolů je vytvoření SW nástrojů pro transformaci a propojování dat. Projekt je spolufinancován Technologickou agenturou České republiky, podpořen Fondem Otakara Motejla a podílí se na něm Vysoká škola ekonomická v Praze a Univerzita Karlova v Praze. [26]

3.5 LOD v zahraničí

Část zahraničních nebo mezinárodních projektů týkajících se otevřených dat byla představena již v kapitole 2.7, v této části jsou představeny projekty zaměřené přímo na propojitelná otevřená data.

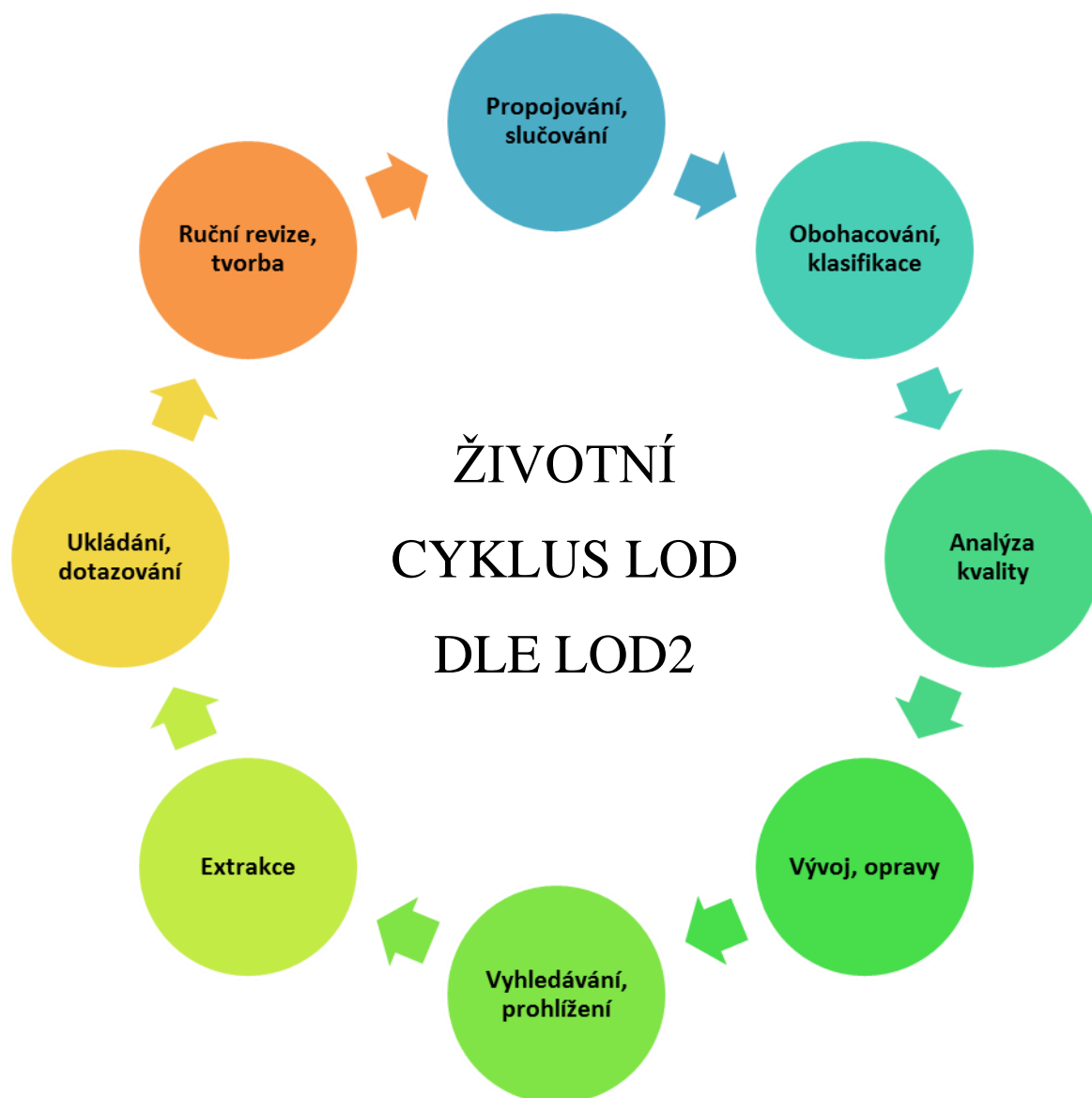
LOD2

Projekt zaměřený na vývoj SW pro otevřená propojitelná data. Tento projekt je spolufinancován Evropskou Komisí. Byl založen v roce 2010 a spojuje výzkumníky, společnosti a poskytovatele služeb z 11 evropských zemí.

Na webu existuje mnoho problémů, které brání vydavatelům i uživatelům v rozvoji LOD.

„Projekt LOD2 se snaží tyto problémy řešit pomocí:

- *podnikově vhodných nástrojů a metodik pro zveřejňování a řízení velmi velkého množství strukturovaných informací na datovém webu,*
- *testbed a bootstrap sítě vysoce kvalitních multi-domén a vícejazyčné ontologie ze zdrojů jako je Wikipedia nebo OpenStreetMap,*
- *algoritmů založených na strojovém učení pro automatické propojování a spojování dat z webu,*
- *adaptivních nástrojů pro vyhledávání, prohlížení a určování vlastnictví propojitelných dat. [27 s. 2]*



Obrázek 19. - Životní cyklus LOD

Zdroj: stack.lod2.eu, vlastní zpracování

COMSODE

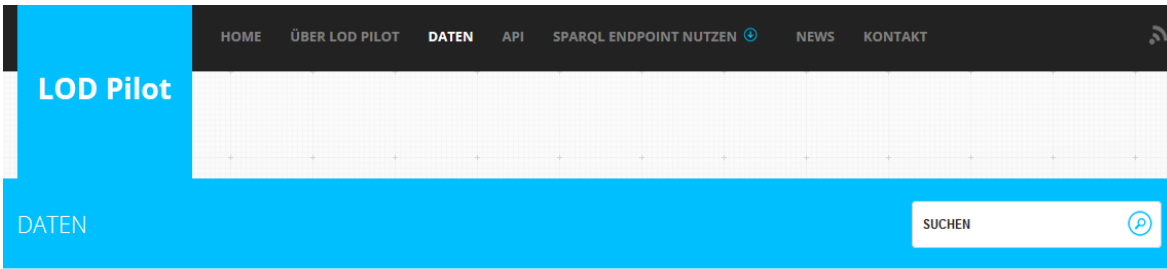
Jedná se o výzkumný a vývojový projekt plánovaný na 24 měsíců (září 2013 až srpen 2015). Podstata tohoto projektu je postavena na faktu, že většina otevřených dat v současnosti postrádá vlastní propojenost a tím je negativně ovlivněna využitelnost těchto dat. Projekt je financován Evropskou unií a jeho hlavním cílem je vytvoření SW nástrojů a metodik, které usnadní proces publikace a využitelnost otevřených dat. Výsledný nástroj by měl umožnit transformaci základních dat do formátů otevřených propojitelných dat, vyčištění dat, jejich anonymizaci, propojení a katalogizaci.

Velká Británie

Již zmíněný britský portál pro otevřená data data.gov.uk umožňuje při vyhledávání použít velké množství různých filtrů, včetně nastavení pouze formátů pro propojitelná otevřená data. Též je zde možné vyhledávat dle hvězdičkové úrovně otevřenosti – 5 hvězdiček pro propojitelná otevřená data.

Rakousko

Rakousko pro otevřená propojitelná data vytvořilo samostatný portál lodpilot.at. Projekt sdružuje data z vládní i regionální úrovně státní správy, ale také od nevládních organizací.



The screenshot shows the LOD Pilot website interface. At the top is a navigation bar with links: HOME, ÜBER LOD PILOT, DATEN (active), API, SPARQL ENDPOINT NUTZEN, NEWS, and KONTAKT. Below the navigation bar is a large blue header with the text 'LOD Pilot' and 'DATEN'. A search bar with the placeholder 'SUCHEN' is located on the right side of the header. Below the header, there is a text block stating: 'Finden Sie hier die Liste der derzeit verfügbaren Daten, die als LOD aufbereitet wurden. Sie können diese via SPARQL-Abfrage jederzeit abrufen.' Below this text is a table with 5 columns: dataset, publisher, identifier, language, and license. The table contains 8 rows of dataset information.

dataset	publisher	identifier	language	license
http://linkeddata.gv.at/statistik-austria/bauwerke/void/dataset/de#stage	http://www.statistik.at	bauwerke	de	https://creativecommons.org/licenses/by/3.0/at/deed.de
http://linkeddata.gv.at/statistik-austria/bauwerke/void/dataset/en#stage	http://www.statistik.at	bauwerke	en	https://creativecommons.org/licenses/by/3.0/at/deed.de
http://linkeddata.gv.at/statistik-austria/berufe/void/dataset/de#stage	http://www.statistik.at	berufe	de	https://creativecommons.org/licenses/by/3.0/at/deed.de
http://linkeddata.gv.at/statistik-austria/funktionelle-klassifikation/void/dataset/de#stage	http://www.statistik.at	funktionelle-klassifikation	de	https://creativecommons.org/licenses/by/3.0/at/deed.de
http://linkeddata.gv.at/statistik-austria/funktionelle-klassifikation/void/dataset/en#stage	http://www.statistik.at	funktionelle-klassifikation	en	https://creativecommons.org/licenses/by/3.0/at/deed.de
http://linkeddata.gv.at/statistik-austria/gueter/void/dataset/de#stage	http://www.statistik.at	gueter	de	https://creativecommons.org/licenses/by/3.0/at/deed.de
http://linkeddata.gv.at/statistik-austria/gueter/void/dataset/en#stage	http://www.statistik.at	gueter	en	https://creativecommons.org/licenses/by/3.0/at/deed.de
http://linkeddata.gv.at/eunis/habitats/void/dataset/en#stage	http://www.eea.europa.eu	habitats	en	https://creativecommons.org/licenses/by/3.0/at/deed.de

Obrázek 20. - Rakouský portál pro otevřená propojitelná data - lodpilot.at

Zdroj: <http://lodpilot.at/daten/>

4 Možnosti využití dat ve státní správě

Při hledání možností využití otevřených dat ve státní správě je nutné si uvědomit, jak na tom v této oblasti vlastně česká státní správa je v porovnání s ostatními zeměmi. Tato kapitola se v závislosti na mezerách ve využití otevřených dat v ČR snaží nastínit jejich lepší využití. Zároveň je zde nastíněn odhad budoucího vývoje.

4.1 Současná pozice české státní správy

Dle indexu pro otevřená data (Open data index), který pravidelně každý rok vydává organizace OKFN se ČR posunula z 30. místa (60 zapojených zemí) v roce 2013 na 12. místo (97 zapojených zemí) v roce 2014.

Rank	Place	Transport Timetables	Government Budget	Government Spending	Election Results	Company Register	National Map	National Statistics	Legislation	Postcodes / Zipcodes	Pollutant Emissions	Score
1	United Kingdom											97%
2	Denmark											83%
3	France											80%
4	Finland											73%
5	Australia											72%
5	New Zealand											72%
7	Norway											71%
8	United States											70%
9	Germany											69%
10	India											68%
11	Taiwan											67%
12	Colombia											66%
12	Czech Republic											66%
12	Sweden											66%
12	Uruguay											66%
16	Iceland											64%

Obrázek 21. - Open data index 2014 - pořadí zemí

Zdroj: <http://index.okfn.org/place/>

Kromě tohoto indexu existují i další podobné žebříčky, např. Open Data Barometr, ve kterém se ČR v roce 2013 umístila na 22. místě a v roce 2014 na 17. místě.

I přes zlepšující se přístup k otevřeným datům jsou v české státní správě stále velké možnosti jejich využití a značné prostory ke zlepšení.

4.2 Problémy otevřených dat v ČR

Před určením potenciálu otevřených dat v ČR je nutné se podívat na problémy, které aktuálně nejvíce brzdí jejich rozvoj. Jako hlavní problémy ztráty konkurenceschopnosti ČR se uvádí následující důvody:

- *neefektivita státní správy,*
- *nedůvěra v instituce,*
- *korupce.* [17 s. 13]

Dle OKFN jsou hlavní nedostatky otevřených dat v ČR následující:

- vládní výdaje (nedostupná veřejně, nedostupná zdarma, nedostupná online, nedostupná ve strojově čitelném formátu, nekompletní datasety, nesplňující podmínky otevřené licence, neaktualizovaná),
- registry firem (neaktualizovaná),
- poštovní směrovací čísla (nesplňující podmínky otevřené licence),
- veřejná doprava (nekompletní datasety, nesplňující podmínky otevřené licence),
- legislativa (nedostupná ve strojově čitelném formátu, nekompletní datasety),
- znečištění (nedostupná ve strojově čitelném formátu, nekompletní datasety, neaktualizovaná).

V rámci iniciativy OGP se ČR zavázala k přijmutí opatření, která by měla nedostatky v oblasti otevřených dat v ČR odstranit, přesto mnoho z těchto problémů přetrvává, z toho je možné usoudit, že zavádění konkrétních opatření je v ČR velmi pomalé.

4.3 Potenciál otevřených dat ve státní správě

Otevřená data mají v ČR do budoucna velký potenciál. Na mnoho velkých mediálních kauz se přišlo jen díky tomu, že si někdo dal práci s analýzou dat, která byla veřejně přístupná, jen ne ve správné formě otevřených dat. V případě, že by všechna zveřejněná data byla v ČR přístupná ve správné formě, mohlo by to veřejnému i soukromému sektoru ušetřit nemalé množství financí.

Ve státní správě je nemalé množství informací dostupné pouze na vyžádání. Pokud by tyto informace nebo alespoň surová data byly dostupné předem, bez nutnosti o ně žádat, ušetřilo by to čas a peníze nejen občanům, ale i příslušným úředníkům.

Přes veškeré problémy, které v současné době brzdí otevřená data v české státní správě, je možné očekávat stále se zvětšující počet institucí, které budou otevřená data zavádět. Nejenom díky snaze všech ze zmíněných projektů, ale snad i díky inspiraci a získaným zkušenostem od institucí, která již otevřená data zavedly.

4.4 Kroky ke zlepšení využitelnosti otevřených dat v ČR

Jaké kroky je nutné učinit ke zlepšení situace otevřených dat v ČR? Jedná se například o:

- sjednocení způsobu publikace otevřených dat,
- vytvoření oficiálního katalogu otevřených dat ČR,
- implementace evropských směrnic,
- vytvoření univerzálních datasetů pro jednotlivé oblasti.

Chlapek aj. v rámci zprávy pro ePSIplatform v roce 2014 uvádějí jako nutné změny:

- Publikace datasetů s vysokou prioritou – je nutné je identifikovat a rozvíjet dle požadavků. Tento přístup by měl být upřednostňován před tím, kdy jsou zveřejňovány bezvýznamné datasety, za účelem vytvoření iluze otevřenosti dané instituce.
- Standardizace napříč institucemi veřejného sektoru – standardizace by měla přinést jednodušší integraci podobných datasetů napříč veřejným sektorem.
- Vytvoření návodů pro identifikaci datasetů – pro instituci nemusí být jednoduché určit, které datasety je vhodné zveřejnit, s tím by jim měly pomoci příslušné návody.
- Vytvoření oficiálního katalogu otevřených dat – jak již bylo zmíněno v předchozích kapitolách, v ČR není v současné době oficiální katalog otevřených dat. Vytvoření takového katalogu by mohlo výrazně zjednodušit vyhledávání konkrétních datasetů.
- Implementace aktuálních evropských směrnic a doporučení – ČR má jako členský stát EU povinnost novelizovat své zákony dle aktuálních nařízení EU.
- Vytvoření komunikační strategie – pro lepší rozvoj otevřených dat je nutné zlepšit komunikaci mezi jednotlivými objekty, již se tato problematika týká. Také by bylo vhodné zlepšit propagaci otevřených dat. [28]

5 Návrh standardizace datových setů

Jak je možné řešit problematiku různorodých způsobů a formátů při zveřejňování otevřených dat? Ve chvíli, kdy jsou datasety tvořeny různými lidmi, je problematické, aby dané datasety měly stejnou strukturu a aby je bylo možné propojovat. Záměrem této části je navrhnout metody pro standardizaci datasetů z oblastí veřejného zájmu, na jejichž základě by bylo možné efektivně porovnávat data z jednotlivých institucí. Tato standardizace zde bude demonstrována na městech a krajích, ale podobný systém by mělo být možné aplikovat i v případě jiných institucí, které disponují společnými znaky.

5.1 Dataset

Dataset je ve své podstatě jakýkoliv soubor obsahující data. Vyjadřuje určitou skupinu dat a většinou odpovídá jedné databázové tabulce. Zpravidla se jedná o ucelenou skupinu spolu souvisejících údajů. Charakteristickým znakem datasetu je, že má vypovídající hodnotu bez ohledu na jiné datasety. Každý dataset je obvykle spravován jedním konkrétním subjektem či osobou. Pravděpodobně nejznámějším příkladem datasetu může být jeden soubor aplikace Microsoft Excel.

5.2 Vhodné oblasti pro datasety

Základním předpokladem pro tvorbu daného datasetu je jeho jednoduchost. V případě složitého datasetu hrozí, že daná instituce si plánované zveřejnění dat prostřednictvím daného datasetu rozmyslí a data buď zveřejní v původním formátu či nezveřejní vůbec. Taktéž je neefektivní zveřejňovat datasety, které nikoho nezajímají. Jaké konkrétní oblasti dat by tedy měly být zvoleny pro standardizované datasety je otázka pro dané instituce, v zásadě však existují následující oblasti ve kterých je vhodné taková data hledat:



Obrázek 22. - Vhodné oblasti pro zveřejnění dat

Zdroj: <http://blog.okfn.org/2013/10/03/defining-open-data/>

V případě měst se tedy může jednat například o granty, smlouvy, poplatky, údaje o obyvatelstvu, atd. Samotné oblasti je poté možné rozdělit na podoblasti nebo skupiny, které tvoří samotné řádky datasetu. V případě grantů by se jednalo o jednotlivé kategorie, tedy např. sport, kultura atd. Sloupce datasetu by poté měly být tvořeny jednotlivými atributy. U grantů se může jednat například o požadovanou velikost grantů a skutečnou velikost schválenou městem.

5.3 Standardizace datových setů

Ve chvíli, kdy mají instituce zvoleny data z vybraných oblastí, je nutné určit, jaká konkrétní data má význam zveřejňovat. V roce 2013 se členové Fóra pro otevřená data a představitelé obcí dohodli na výměně zkušeností s otevřenými daty. Společně by měli pracovat na standardizaci procesů, schémat a formátů pro vybrané datasety. Díky tomu by mělo být možné porovnávat datasety ze stejné nebo podobné oblasti z různých měst ČR. [28]

Pro zjednodušení bude pro ukázkou standardizace zvolena jednoduchá tabulka. Převod do vhodnějšího formátu RDF je popsán níže. V rámci samotné standardizace není zvolený formát zas až tak významný. V jakém formátu by se nakonec standardizované tabulky publikovaly, závisí na dohodě příslušných autorit.

Pro ukázkou standardizace bude využito otevřených dat města Opavy a Přerova. Byla vybrána oblast grantů za rok 2015. Město Opava zveřejnilo své granty v jednom xls souboru, kde jsou jednotlivé oblasti rozděleny do listů. Granty jsou zde rozděleny do 3 základních kategorií – sport, kultura, ŽP a EVVO. U města Přerova jsou jednotlivé kategorie grantů jako samostatné pdf soubory a to sice – sport, kultura, sociální, volný čas, zdravotnictví. Nabízí se tedy kategorie z obou měst spojit. Výsledná tabulka za celé kategorie pro obě města zároveň by tedy vypadala takto:

Tab. 3. - Srovnání grantů v Přerově a Opavě

	Přerov	Opava
Sport	2 354 000	1 733 000
Kultura	1 105 000	1 673 000
ŽP a EVVO	0	201 700
Sociální	1 067 000	0
Volný čas	250 000	0
Zdravotnictví	27 000	0

Zdroj: weby měst Přerova a Opavy, vlastní zpracování

Jak je patrné z tabulky, kde jsou uvedena obě města zároveň, je následná standardizace již zveřejněných dat značně neefektivní. Každé město uděluje granty v jiné oblasti, stejně tak nemusí mít veškerá data zveřejněná, některé podobné granty mohou být v každém městě zveřejněny v jiné kategorii a je tedy možné mít u jednotlivých kategorií nuly u některých měst. Podobná situace je i v jiných oblastech. V případě většího množství sloupců by vypovídací hodnota byla ještě menší. Vyplývá z toho, že pro efektivní standardizaci je nutné, aby konkrétní sloupce a řádky byly domluveny již před zveřejněním dat. Konkrétní

sloupce a řádky standardizovaného datasetu by měly být vybrány pečlivě s výhledem na budoucí data.

5.4 Formát dat

Důležité je si říci, v jaké formě by bylo vhodné data prezentovat. Z hlediska dalšího možného využití, případně propojení dat je nutné, aby zvolený formát dat byl strojově čitelný. Z předchozích kapitol vyplývá, že nejvhodnějším formátem by byl formát RDF. Je tedy nezbytné tento formát představit blíže.

„At the simplest level, the Resource Description Framework (RDF) is an XML-based language to describe resources.“ [29 s. 85]

Zjednodušeno přeloženo: v nejjednodušší podobě je RDF jazyk založený na XML, sloužící k popisu zdrojů.

RDF je formát/model čitelný jak pro lidi, tak pro stroje. Model RDF je hlavním nástrojem sémantického webu, jeho základem je popisování zdrojů na webu. Může jím být popsán jakýkoliv zdroj, který je možné identifikovat pomocí URI, např. webová stránka či PDF dokument, ale i entita reálného světa. RDF je sám o sobě abstraktní formát a k jeho zapsání je tak nutný již zmíněný značkovací jazyk XML. [30]

RDF je postaven na pojmu známém jako ontologie. Ontologie je model, který se snaží popsat svět pomocí typů, vlastností a vztahů. Datům je možné pomocí ontologie přiřadit význam. Základním předpokladem je snaha o co nejvěrnější popis reálného světa. Pro konkrétní popisy jsou vytvořeny slovníky. Mezi známé slovníky patří např.: FOAF – slouží k popisu lidí, jejich aktivit nebo vztahů, nebo vCARD – slouží pro popis lidí a organizací, simuluje klasické vizitky. Kromě těchto slovníků může tvůrce samozřejmě vytvářet i vlastní popisy a třídy (s vlastním URL).

Základní vlastností RDF je vytvoření výrazu ve tvaru tzv. RDF trojice: subjekt/podmět-predikát/vlastnost-objekt/předmět. Subjekt je obvykle reprezentován pomocí URL entity (URL uzlu sítě), predikát je vlastnost týkající se entity a objekt je hodnota vlastnosti. Např. ve větě „Petr má rád rohlíky“ je „Petr“ subjekt, „má rád“ predikát a „rohlíky“ jsou objekt. Díky tomu může daný vztah pochopit i stroj. Uživatel, který neví, co má Petr rád se pak může pomocí SPARQL dotazu zeptat stroje: „Co má Petr rád?“ a stroj mu bude schopen odpovědět, že se jedná o rohlíky. Proto, aby počítač věděl, o jakého konkrétního Petra se jedná, je nutné ho jednoznačně identifikovat pomocí URI. Pro stroj tedy daný vztah Petr – má rád – rohlíky vypadá v URL (podmnožina URI) např. takto: <http://petrove.cz/petr5> – <http://vztahy.cz/marad> – <http://pecivo/rohliky>. Ve chvíli, kdy jsou takto propojena data na celém webu, má uživatel možnost klást zajímavé otázky a dostat na ně přímé odpovědi. Využívá toho například společnost Google, kdy je možné se na vyhledávači Google zeptat např. „Kolik je 5 metrů ve stopách?“ a uživateli se dostane přímé odpovědi „16,4041995 stopy“, bez nutnosti dalšího hledání. Z toho vyplývá, že existuje jasně definovaný vztah mezi měrnými jednotkami. Stejně tak je možné takové vztahy definovat i v méně zřejmých případech.

5.5 Práce s RDF

Pro práci s formátem RDF existuje celá řada nástrojů. Každý z těchto nástrojů umožňuje jinou činnost. Některé z nástrojů, které umožňují práci s RDF, byly primárně vytvořeny za jiným, širším účelem, jiné mají naopak velmi úzké využití. Celý proces práce s RDF je možné rozložit do tří hlavních činností: vytvoření RDF, uložení a prohlížení.

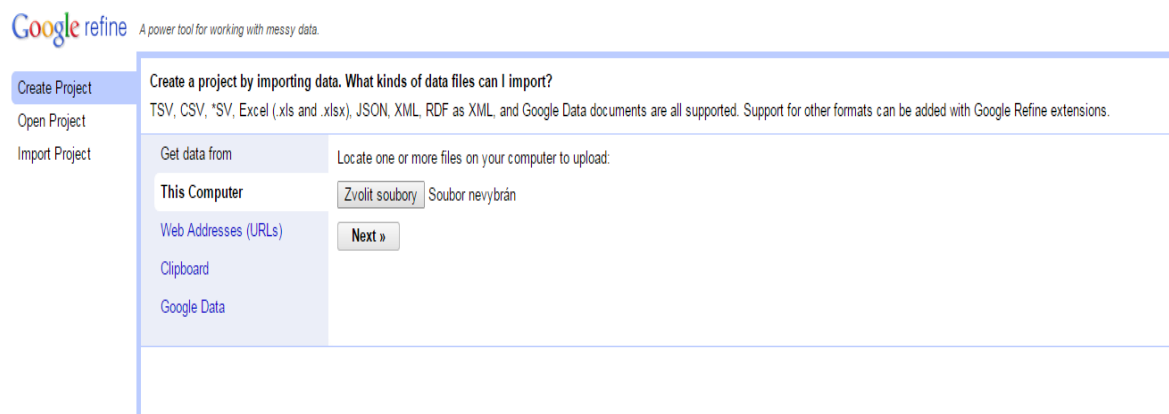
5.5.1 Převod do RDF

Většina dat je k dispozici v jiném formátu než RDF. Jak tedy zajistit aby data byla ve formátu RDF? K tomuto účelu vzniklo mnoho nástrojů, které jsou schopny převést data ze zdrojového formátu do formátu RDF. Neboli na vstupu jsou např. soubory formátů csv, xml, xls či jiné a na výstupu je RDF soubor. Tomuto procesu se říká triplifikace. Kromě propojení dat může převod do RDF znamenat i vyčištění dat. Díky sjednocení

pojmenování skupin pro propojení je možné se zbavit chyb typu přebytečné mezery a podobně. Na webu mezinárodního konsorcia W3, které stojí za vývojem standardů pro WWW, je k dispozici seznam přibližně 250 nástrojů pro práci se sémantickým webem. Mezi ty nejznámější patří: Virtuoso Sponger, Aperture, Any23, OpenRefine, Alchemy, Tripliser a další.

Jedním z populárních nástrojů je již zmíněný OpenRefine. OpenRefine (dříve Google Refine) nabízí velké možnosti v oblasti čištění a úpravy dat. Jednou z možností, které nabízí, je právě převod dat do RDF. Jedná se o open source projekt a samotný program je možné si stáhnout zdarma z webu. Ke správné funkcionalitě s RDF je třeba přidat rozšíření RDF Refine.

Po stažení stačí daný soubor rozbalit a následně otevřít soubor s názvem google-refine. Otevře se zároveň příkazový řádek s lokálním virtuálním serverem a aplikace v prohlížeči. Nejdříve je nutné vytvořit nový projekt. Do programu je možné data nahrát z různých zdrojů – z webové adresy, z počítače, či přímo překopírovat.



Obrázek 23. - OpenRefine - nahrání souboru s daty

Zdroj: OpenRefine, vlastní zpracování

Při nahrání souboru z počítače je možné zvolit např. xls soubor. V tomto případě tabulka s kraji ČR, počty jejich obyvatel, průměrnou mzdou a nezaměstnaností.

The screenshot shows the OpenRefine web interface. At the top, there's a 'Google refine' logo and a 'Kraje' tab. Below the logo, there's a 'Facet / Filter' button and an 'Undo / Redo' button. The main area displays a table with 14 rows. The table has five columns: 'All', 'Kraj', 'Počet obyvatel', 'Hrubá mzda', and 'Nezaměstnanos'. The data is sorted by 'Kraj' and shows the following information:

	Kraj	Počet obyvatel	Hrubá mzda	Nezaměstnanos
1.	Hlavní město Praha	1252202	32910	0.0508
2.	Jihočeský kraj	637143	22591	0.0667
3.	Jihomoravský kraj	1171158	24329	0.0842
4.	Karlovarský kraj	299445	21553	0.0843
5.	Kraj Vysočina (Jihlavský)	510099	22706	0.076
6.	Královéhradecký kraj	551689	22842	0.0645
7.	Liberecký kraj	436767	23163	0.0777
8.	Moravskoslezský kraj	1216732	22964	0.0996
9.	Olomoucký kraj	636015	22284	0.092
10.	Pardubický kraj	516315	22193	0.0853

Obrázek 24. - OpenRefine - tabulka

Zdroj: OpenRefine, vlastní zpracování

Následně je možné např. přidávat slovníky, ze kterých můžeme vyhledávat predikáty, propojovat subjekty a objekty s jinými datasety z různých databází, atd.

5.5.2 Uložení RDF

Vytvořený RDF dokument je možné buď uložit do databáze nebo serializovat. Takové databáze jsou pak buď postaveny na klasických serverech typu MySQL, Oracle, atd. nebo jsou postaveny přímo pro RDF. Takové databázi se říká Triplestore a je optimalizovaná přímo pro RDF trojice. Mezi známé databáze patří např.: DBpedia, Jena TDB, Europeana, OpenLink Virtuoso a další.

RDF se vyskytuje v několika serializacích (verze zápisu/uložení). Mezi nejznámější patří: RDF/XML, N3, NTriples, Turtle, TriX, TriG nebo RDFa. Nejoblíbenější z nich je Turtle, využívající příponu názvu souboru *.ttl. Serializace RDF jsou vzájemně převoditelné. Existuje několik nástrojů pro převod mezi serializacemi, např.: Any23 nebo Raptor RDF parser. Serializací RDF vznikne v podstatě jednoduchý textový soubor. Serializace je tudíž nevhodná hlavně v případě souborů s velkým objemem dat uložených ve trojicích. V takovém případě je efektivnější využít uložení do databáze, která může díky indexování rozhodnout, která data načíst do paměti a výrazně tak zkrátit dobu vyhledávání.

V nástroji OpenRefine je možné vytvořený RDF soubor vyexportovat např. do Turtle serializace a v této podobě následně uložit (v případě předchozí ukázky jako Kraje.ttl):

```
@prefix : <> .
@prefix dbpedia-owl: <http://dbpedia.org/ontology/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<http://localhost:3333/0> dbpedia-owl:region "Hlavní město Praha" ;
    owl:numberofinhabitants "1252202.0" ;
    dbpedia-owl:salary "32910.0" ;
    owl:percentageofunemployment "0.0508" .

<http://localhost:3333/1> dbpedia-owl:region "Jihočeský kraj" ;
    owl:numberofinhabitants "637143.0" ;
    dbpedia-owl:salary "22591.0" ;
    owl:percentageofunemployment "0.0667" .

<http://localhost:3333/2> dbpedia-owl:region "Jihomoravský kraj" ;
    owl:numberofinhabitants "1171158.0" ;
    dbpedia-owl:salary "24329.0" ;
    owl:percentageofunemployment "0.0842" .

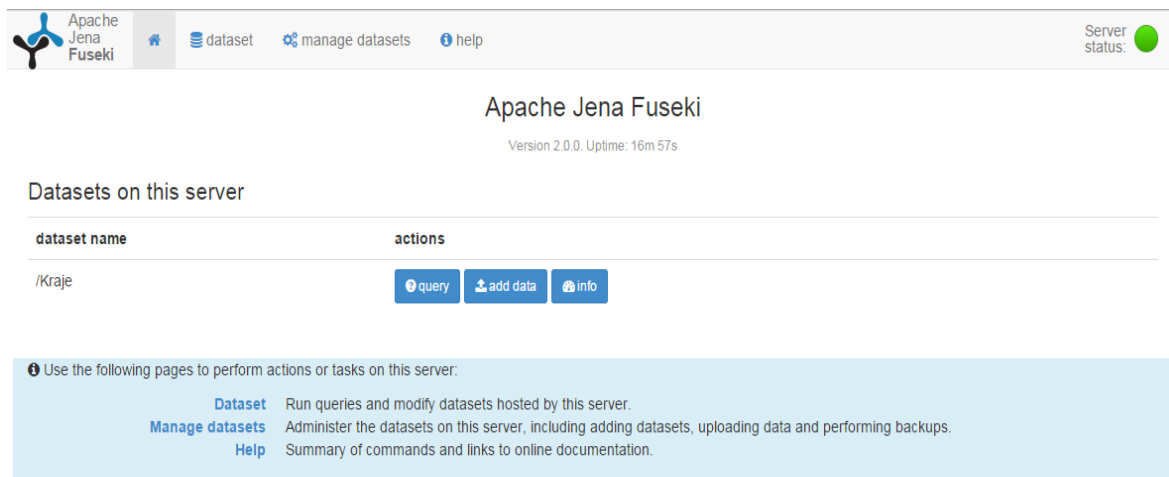
<http://localhost:3333/3> dbpedia-owl:region "Karlovarský kraj" ;
    owl:numberofinhabitants "299445.0" ;
    dbpedia-owl:salary "21553.0" ;
    owl:percentageofunemployment "0.0843" .

<http://localhost:3333/4> dbpedia-owl:region "Kraj Vysočina (Jihlavský)" ;
    owl:numberofinhabitants "510099.0" ;
    dbpedia-owl:salary "22706.0" ;
    owl:percentageofunemployment "0.076" .
```

Obrázek 25. - Turtle syntaxe

Zdroj: OpenRefine, vlastní zpracování

Pro uložení do databáze je možné využít například nástroj od společnosti Apache Jena Fuseki, který umožní vytvoření virtuálního serveru na lokálním počítači. Stejně jako OpenRefine je tento SW open source a je možné si ho volně stáhnout. Po stažení je opět nutné daný soubor rozbalit a po rozbalení spustit soubor fuseki-server. Jena Fuseki server se poté sám spustí na portu 3030. K tomu je možné přistoupit přes webový prohlížeč po zadání adresy localhost:3030. Na server je poté možné nahrát dříve vytvořený soubor Kraje.ttl.



Obrázek 26. - Apache Jena Fuseki - nahrání datasetu

Zdroj: Apache Jena Fuseki, vlastní zpracování

5.5.3 Prohlížení RDF

Pro následnou inferenci nad RDF daty se využívá jazyk SPARQL, který umožňuje vytvářet dotazy nad triplestore databázemi. Jak je z názvu patrné, má podobnou syntaxi jako SQL. Díky přítomnosti URI jsou takové dotazy mnohem efektivnější než full-textové vyhledávání.

V závislosti na způsobu uložení RDF existují v zásadě dvě možnosti, jak se nad takovými daty dotazovat. V případě uložení do serializace je možné využít některý z nástrojů pro vytváření SPARQL dotazů. Z pohledu otevřených dat ovšem dává větší význam tvorba dotazů nad některou z veřejně dostupných databází. K tomu se využívá tzv. SPARQL endpoint, což je zjednodušeně webová služba, která nám pomocí SPARQL dotazů umožňuje klást konkrétní otázky nad danou databází. Jedním ze známých nástrojů pro dotazování nad daty je např. Virtuoso SPARQL Query Editor.

SPARQL dotaz je možné vznést i v případě ukázky vytvořeného lokálního serveru v prostředí Apache Jena Fuseki přes položku query:

The screenshot shows the Apache Jena Fuseki web interface. At the top, there is a navigation bar with links for 'query', 'upload files', 'edit', and 'info'. Below this, the 'SPARQL query' section is active. It includes a text area for entering a query, with a hint: 'To try out some SPARQL queries against the selected dataset, enter your query here.' Below the text area, there are two buttons: 'Selection of triples' and 'Selection of classes'. Further down, there is a 'PREFIXES' section with buttons for 'rdf', 'rdfs', 'owl', and 'xsd'. Below the prefixes, there are three dropdown menus: 'SPARQL ENDPOINT' (set to 'http://localhost:3030/Kraje/sparql'), 'CONTENT TYPE (SELECT)' (set to 'JSON'), and 'CONTENT TYPE (GRAPH)' (set to 'Turtle'). At the bottom, there is a large text area for the query, containing the following SPARQL query:

```
1
2
3 SELECT ?subject ?predicate ?object
4 WHERE {
5   ?subject ?predicate ?object
6 }
7 LIMIT 25
```

On the right side of the query text area, there are three icons: a share icon, a full screen icon, and a play icon.

Obrázek 27. - Apache Jena Fuseki - SPARQL dotazování

Zdroj: Apache Jena Fuseki, vlastní zpracování

5.6 Navrhovaný proces pro zavedení standardizace

K tomu, aby měla standardizace smysl, je nutná shoda všech stran, které by se na ní měly podílet. Jaký by tedy měl být postup vedoucí ke standardizaci?

1. Úvodní schůzka zástupců vybraných měst a odborníků v oblasti otevřených dat. Na této schůzce by se měly dohodnout tyto body:
 - a. Zvolení oblastí pro zveřejnění. Nutné vybírat data vhodná pro zveřejnění – jak po faktické, tak i právní stránce.
 - b. Zvolení konkrétních sloupců a řádků pro standardizaci mezi městy. Jedná se o klíčový bod a je třeba zvážit veškeré atributy, které budoucí zveřejňovaná data budou mít.
 - c. Výběr formátů v jakém data zveřejnit. V ideálním případě ve formátu RDF, u dat, kde to není výhodné, je nutné zvolit jiný vhodný formát. Měla by být zachována snaha o vybrání strojově čitelného formátu.
 - d. Jednotný web/aplikace pro zveřejnění vybraných dat. Pro uživatele by bylo nejvhodnější, aby standardizovaná data byla dostupná na jednom místě. Je nutné učinit rozhodnutí, zda zvolit jednoduchý web na kterém budou uživatelé stahovat jednotlivé soubory nebo vyvinutí webové aplikace a stažení dat použít pouze jako možnost v rámci aplikace. Zde je též nutné pro budoucí uživatele stanovit podmínky užití.
 - e. Jakým způsobem mohou města zvolená data zveřejnit? Kromě způsobu publikace je taktéž nutné určit, kdo bude publikaci zajišťovat.
 - i. Pomocí vlastních zdrojů.
 - ii. Pomocí specializovaných IT firem – dodávky dat daným firmám, které se postarají o zveřejnění.
 - iii. Vyvinutí SW nástrojů – IT společnost vyvine SW, který budou daná města sama využívat.
 - f. Katalogizace. Pokud by již byl v této době potenciálně vytvořen oficiální katalog dat ČR, měla by se zveřejněná data katalogizovat.
 - g. Propagace. Aby měl celý proces smysl, je nutné o zveřejněných datech informovat veřejnost.

- h. Aktualizace. V neposlední řadě je třeba se dohodnout na způsobu aktualizace. Tu je možné provádět několika způsoby:
 - i. Metoda přepisování – údaje jednou zveřejněné budou vymazány a místo nich budou nahrány nové údaje.
 - ii. Metoda přidávání – k již zveřejněným datům budou přidána nová data.
 - iii. Metoda dotazování – veškeré údaje jsou uloženy v databázi a uživatel k nim přistupuje pomocí dotazů.
- i. Kontrola. Bylo by vhodné vytvořit mechanismy pro kontrolu zveřejňovaných dat. Zda například nedochází ke zbytečnému zdvojování zveřejňovaných dat a tím pádem i ke zbytečným výdajům. Takovou kontrolu je možné provádět např.:
 - i. Vytvořením speciálního týmu, který by se zabýval problematikou kontroly dat zveřejňovaných v rámci standardizovaných datasetů.
 - ii. Umožněním kontroly občanům. Pokud by se plány, aktuální stavy a další informace týkající se prací na standardizovaných datasetech zveřejňovaly na internetu, mohli by kontrolu provádět sami občané.
- 2. Pravidelné status meetingy vybraných zástupců. V oblasti, která se neustále mění je nezbytná dohoda na pořádání pravidelných jednání pro zhodnocení stavu a případné změny.

Je jasné, že ne veškerá data by byla vhodná ke zveřejnění tímto standardizovaným systémem. Některá data jsou specifická pouze pro dané město a není je tak možné použít pro standardizaci mezi městy. Neznačená to ale, že by se města měla vzdát myšlenky tato data zveřejnit.

6 Vyhodnocení návrhu

Tab. 4 - Vyhodnocení návrhů

	Publikování v RDF	Standardizace datasetů	Proces zavedení standardizace
Finanční náročnost	Finančně nenáročné. V závislosti na nakoupeném SW/sluzbách.	Finančně nenáročné. V závislosti na nakoupeném SW/sluzbách.	Finančně nenáročné. Náklady na příslušné množství FTE během jednání.
Technická náročnost	Technicky mírně náročnější, nutná dobrá znalost IT.	Může vzrůst i poklesnout, v závislosti na tom, co město zveřejňovalo před standardizací.	V závislosti na množství participujících stran.
Nutné změny	Proškolení IT specialistů, najmutí externího dodavatele.	Nenáročné u měst, která již některá z dat publikovala před standardizací.	Přijmutí procesu standardizace, včetně následných pravidelných schůzek
Výhody	Propojitelnost dat.	Možnost efektivního porovnávání mezi městy.	Možnost vytvoření dlouhodobější dohody.
Nevýhody	Technická náročnost může některé instituce odradit. Není vhodné pro všechna data.	Možný negativní efekt v případě špatně nastavené standardizace.	Nutná dohoda více stran.
Reálnost	Menší. Náročné.	Větší. Na užitečnosti zavedení standardizace panuje všeobecná shoda.	Střední. Proces je variabilní.

Zdroj: vlastní zpracování

6.1 Nároky na implementaci

Nároky na implementaci, ať už v oblasti metodik a procesů či konkrétních standardizovaných datasetů, nejsou v jednotlivých institucích nijak výrazné. Dle zkušeností z institucí, kde již svá data otevřely, je časová náročnost samotné implementace odhadem několik dnů či týdnů. Příprava k implementaci může trvat déle v závislosti na velikosti instituce a objemu zveřejňovaných dat.

6.2 Výhody standardizace datasetů

Jaké výhody by vlastně standardizace popsaná v předchozí kapitole přinesla? Pro samotná města by standardizace mohla zpočátku přinést určité množství zvýšených nákladů. V celkovém důsledku by takové řešení přineslo finanční úspory díky možnosti efektivního porovnání. V takovém porovnání s obdobnou institucí by mohly například vyjít najevo informace o neefektivitě instituce v určité oblasti, na kterou by se pak daná instituce mohla zaměřit. Výhody standardizace datasetů jsou ale především celospolečenské. Díky možnosti efektivního porovnání je možné nalézt neefektivitu, která často stojí nemalé množství finančních prostředků.

Závěr

V českém veřejném prostoru a v jednotlivých institucích se objevuje celá řada výmluv, proč právě u nich není vhodné data zveřejňovat. Data, respektive informace jsou v dnešní době nástrojem moci a politici se moci neradi vzdávají. Pokud by však úřady využívaly otevřená data, mohly by ušetřit čas a především peníze nejen sobě, ale i občanům.

Jedním z cílů této práce bylo osvětlit čtenáři situaci otevřených dat v ČR i ve světě. Toto bylo provedeno analýzou v druhé kapitole. Čtenář by po prostudování této kapitoly měl mít základní přehled o organizacích a projektech v ČR i ve světě. Byly též uvedeny důležité legislativní aspekty problematiky otevřených dat. Vlastní kapitola byla věnována problematice propojitelných otevřených dat, u kterých byly také zmapovány klíčové organizace a projekty, jak v ČR, tak i ve světě.

Druhým cílem v této práci bylo definování potenciálu otevřených dat ve státní správě. Na základě identifikace nedostatků v české státní správě byla stanovena doporučení a kroky vedoucí ke zlepšení využitelnosti otevřených dat v ČR.

Posledním cílem bylo navrhnout vhodné postupy a metody vedoucí ke standardizaci datasetů. Navrhovaná řešení standardizace si kladou za cíl specifikovat, zefektivnit a zkvalitnit proces publikace otevřených dat. Zároveň by tato řešení měla přinést větší využitelnost otevřených dat.

Otevřená data jsou stále poměrně mladou oblastí, ve které dochází k neustálým změnám. Autor musel již během vypracovávání této práce několikrát uvedené informace aktualizovat. Přesto byl kladen důraz na to, aby veškeré informace zde uvedené, byly co možná nejaktuálnější.

I přes fakt, že většina otevřených dat v ČR není v současné době dostupná v nejlepší možné formě otevřených propojitelných dat, je nutné konstatovat, že situace se zlepšuje. Data zveřejněná alespoň jako jednoduchá otevřená data, například ve formě Excelové tabulky, může kdokoli využít a zpracovat tak, že získají mnohem větší význam. Občan,

který má dostatek informací, se může lépe rozhodovat a zvýšit tak konkurenceschopnost celé země. Česká republika má ambice řadit se mezi moderní a vyspělé země s efektivní a fungující demokracií, otevřená data jsou jednou z možností, jak se tomuto cíli přiblížit.

Seznam použité literatury

Citace

- [1] SKLENÁK, Vilém aj. *Data, informace, znalosti a Internet*. 1. vyd. Praha: C. H. Beck, 2001. ISBN 80-7179-409-0
- [2] GEIGER, C. P. and J. VON LUCKE. *Open Government and Linked Open Government Data*. [online]. 2012, [vid. 2014-11-10]. Dostupné z: <http://www.jedem.org/article/view/143/115>
- [3] BOČEK J., M. BERG a P. BOUCHAL. *Otevřená data ve státní správě: Nová éra rozhodování*. Praha: Nadace Open Society Fund a Otevřená společnost, o.p.s., 2012. ISBN 978-80-871110-24-9
- [4] Ministerstvo vnitra České republiky. *Otevřená data*. [online]. [vid. 2014-11-15]. Dostupné z: <http://www.mvcr.cz/clanek/otevrena-data.aspx>
- [5] OCHRANA F., J. PAVEL a L. VÍTEK. *Veřejný sektor a veřejné finance*. Praha: Grada, 2010. 264 s. ISBN 978-80-247-3228-2
- [6] LIDINSKÝ V. *eGovernment bezpečně*. 1. vyd., Praha: Grada, 2008. ISBN 978-80-247-2462-1
- [7] Digital standards organization. *Definition of a Free and Open Standard*. [online]. [vid. 2015-04-05]. Dostupné z: <http://www.digistan.org/open-standard:definition>
- [8] HUIJBOOM, N. and T. VAN DEN BROEK. *Open data: an international comparison of strategies*. European Journal of ePractice. 12. ed. [online]. 2011, [vid. 2015-01-30]. ISSN 1988-625X. Dostupné z: <http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012%201.pdf>
- [9] KUČERA, J. *Tutoriál: Otevírání a propojování dat*. [online]. Praha: Vysoká škola ekonomická v Praze, 2014. Dostupné z: http://www.cssi.cz/cssi/system/files/all/Seminar_CSSI_6-6-2014_tutorial%20dopoledne.zip
- [10] CHLAPEK, D., J. KUČERA a M. NEČASKÝ. *Metodika publikace otevřených dat veřejné správy ČR: verze 1.0*. In: *Boj s korupcí v České republice* [online]. 2011, [vid. 2014-10-08]. Dostupné z: <http://www.korupce.cz/assets/dokumenty/aktuality/Metodika-publikace-otevrenych-dat.pdf>
- [11] MYNARZ J. *Katalog dat České republiky* [online]. [vid. 2015-04-03]. Dostupné z: <http://www.datablog.cz/clanky/katalog-dat-ceske-republiky>

- [12] Soutěž Otevřeno x Zavřeno 2014. Praha: Otevřená společnost, 2014. Dostupné z: <http://www.mmdecin.cz/obcan/zpravodajstvi/zpravy-z-radnice/item/1068-decin-nominovan-jako-otevreny-urad-roku-hlasovani-verejnosti>
- [13] Statutární město Děčín. *Otevřená data*. [online]. Děčín, [vid. 2015-03-05]. Dostupné z: http://www.mmdecin.cz/dokumenty/cat_view/238-otevrena-data
- [14] Česko. Zákon č. 106/1999 sb. ze dne 11. května 1999 o svobodném přístupu k informacím. In: *Sbírka zákonů České republiky*. 1999. ISSN 1211-1244. Dostupné z: <http://www.mkcr.cz/scripts/detail.php?id=325>
- [15] MANYIKA, J., ed. *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute, 2011. ISBN 978-0983179696
- [16] GALBAVÝ, Luděk. *Proč otevřená data a jaký mají přínos pro veřejnou správu?*. [online]. Svaz měst a obcí České republiky, [vid. 2015-04-22]. Dostupné z: <http://www.smocr.cz/cz/oblasti-cinnosti/informatika/proc-otevrena-data-a-jaky-maji-prinos-pro-verejnou-spravu.aspx>
- [17] BOČEK J., J. MRÁČEK a J. MYNARZ. *Otevřená data: Příležitost pro Českou republiku*. Praha: Nadace Open Society Fund, 2012. ISBN 8087725026
- [18] DEMERSSEMAN W. *Dissed Trust: America's Crisis of Truth, Faith, and Freedom*. 1th ed., Bloomington: WestBow, 2010. ISBN 978-1-44970-367-7.
- [19] Open Knowledge Foundation. *About*. [online]. [vid. 2015-02-15]. Dostupné z: <https://okfn.org/about/>
- [20] KU Leuven. *The LAPSI 2.0 project overview*. [online]. Leuven, [vid. 2015-03-10]. Dostupné z: <http://www.lapsi-project.eu/lapsi-20-project-overview>
- [21] BERG, Michal. *Otevřená data a jejich byznys modely: kde v nich hledat peníze?* [online]. Datablog.cz, [vid. 2015-03-15]. Dostupné z: <http://www.datablog.cz/clanky/byznys-modely>
- [22] HARVEY, Lewis. *Governments looking for economic ROI must focus on open data with business value*. [online]. Radar, [vid. 2015-03-20]. Dostupné z: <http://radar.oreilly.com/2013/02/roi-open-data-economy-value.html>
- [23] LIPTON, Peter. *Inference to the Best Explanation*. 2. ed. London: Routledge, 2004. 232 p. ISBN 978-0-415-24203-5
- [24] WOODCOCK, J., ed. *Slovník výpočetní techniky: Výklad standardních pojmů pro vědu, školství a obchod*. 1. ed. Praha: Plus, 1993. 421 p. ISBN 8085297485
- [25] THURNER, Thomas. *The Stack*. [online]. LOD2, [vid. 2015-04-10]. Dostupné z: <http://stack.lod2.eu/blog/>

- [26] CHLAPEK, Dušan. *Otevřená propojitelná data v oblasti veřejných rozpočtů*. [online]. Otevřená data na VŠE, [vid. 2015-05-01]. Dostupné z: <http://opendata.vse.cz/tacr/mf/>
- [27] AUER, S., V. BRYL and S. TRAMP. *Linked Open Data - Creating Knowledge Out of Interlinked Data*. 1st ed. Springer International Publishing, 2014, 215 p. ISBN 978-3-319-09846-3
- [28] CHLAPEK, D., J. KUČERA a M. NEČASKÝ. *Open data and PSI in the Czech Republic*. Praha: European Public Sector Information Platform, 2014, 25 p. Dostupné z: <http://www.epsiplatform.eu/content/open-data-and-psi-czech-republic>
- [29] DACONTA, Michael C., ed. *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. 1st ed. Wiley, 2003, 312 p. ISBN 978-0471432579
- [30] MLÝNKOVÁ I. aj. *XML technologie: Principy a aplikace v praxi*. Praha: Grada Publishing, 2008, 272 s. ISBN 978-80-247-2725-7

Bibliografie

MRÁČEK J., J. ČEPICKÝ a J. BOČEK. Jak otevírat data? [online]. Praha: Fond Otakara Motejla, 2014. [vid. 2014-06-08]. ISBN 978-80-87725-15-3.

Dostupné z: <http://www.otevrenadata.cz/res/data/001/003498.pdf>

GURIN, J. Open Data Now: The Secret to Hot Startups, Smart Investing, Sawy Marketing and Fast Innovation. 1.vyd. New York: McGraw-Hill, 2014. ISBN 0071829784

BOČEK J., J. MRÁČEK a J. MYNARZ. Otevřená data: Příležitost pro Českou republiku. Praha: Nadace Open Society Fund, 2012. ISBN 8087725026

KUČERA, Jan, CHLAPEK, Dušan. Benefits and Risks of Open Government Data. Journal of systems integration [online]. Praha: Vysoká škola ekonomická v Praze, 2014. [vid. 2014-08-08]. ISSN 1804-2724.

Dostupné z: <http://www.si-journal.org/index.php/JSI/article/viewFile/185/136>

MUSTAFA M. Corporate Use of Open Government Data. [online]. Vienna: Technical University of Vienna, 2012. [vid. 2014-10-05].

Dostupné z: <http://cdn.open3.at/wp-content/uploads/2012/06/MM-OGD-JH-Final-v2.pdf>

Elektronická databáze článků ProQuest (knihovna.tul.cz).

Seznam příloh

Příloha A **Formáty dat**

Příloha B **Přehled 10 nejlepších aplikací v ČR pro otevřená data 2014**

Příloha C **Ukázky z datasetů**

Příloha A Formáty dat

V příloze A je k dispozici popis nejčastějších formátů využívaných pro otevřená data.

Tab. A1 - Popis nejčastějších formátů využívaných pro otevřená data.

Formát	Popis formátu
PDF	PDF (Portable Document Format) je formát souboru používaný k prezentaci a výměně dokumentů. Je nezávislý na softwaru, hardwaru nebo operačním systému. Formát PDF, který vyvinula společnost Adobe, je nyní otevřený standard pod záštitou Mezinárodní organizace pro normalizaci (ISO). Soubory PDF mohou obsahovat odkazy a tlačítka, pole formulářů, zvuk, video a obchodní logiku. Mohou obsahovat také elektronický podpis a je možné je snadno zobrazit pomocí softwaru Acrobat Reader DC.
DOC(X)	Základní formát programu Microsoft Word společnosti Microsoft. Ve starších verzích DOC, v novějších DOCX.
TXT	Jedná se o soubor obsahující pouze textová data. Nejznámějším programem pro práci s textovým souborem je poznámkový blok ve Windows.
HTML	HyperText Markup Language (zkratka HTML) je název značkovacího jazyka používaného pro tvorbu webových stránek, které jsou propojeny hypertextovými odkazy. HTML je hlavním z jazyků pro vytváření stránek v systému World Wide Web, který umožňuje publikaci dokumentů na Internetu.
XLS(X)	XLS je přípona souborů vytvořených v aplikaci Microsoft Excel. Data jsou uložena v binárním formátu. Jedná se o předchůdce formátu XLSX.
CSV	CSV (Comma-separated values, hodnoty oddělené čárkami) je jednoduchý souborový formát určený pro výměnu tabulkových dat. Soubor ve formátu CSV sestává z řádků, ve kterých jsou jednotlivé položky odděleny znakem čárka (.). Hodnoty položek mohou být uzavřeny do uvozovek ("), což umožňuje, aby text položky obsahoval čárku. Pokud text položky obsahuje uvozovky, jsou tyto zdvojeny.

JSON	JavaScript Object Notation (JavaScriptový objektový zápis, JSON) je způsob zápisu dat (datový formát) nezávislý na počítačové platformě, určený pro přenos dat, která mohou být organizována v polích nebo agregována v objektech. Vstupem je libovolná datová struktura (číslo, řetězec, boolean, objekt nebo z nich složené pole), výstupem je vždy řetězec. Složitost hierarchie vstupní proměnné není teoreticky nijak omezena.
XML	Extensible Markup Language (zkráceně XML, česky rozšiřitelný značkovací jazyk) je obecný značkovací jazyk, který byl vyvinut a standardizován konsorciem W3C. Je zjednodušenou podobou staršího jazyka SGML. Umožňuje snadné vytváření konkrétních značkovacích jazyků (tzv. aplikací) pro různé účely a různé typy dat. Používá se pro serializaci dat, v čemž soupeří např. s JSON či YAML. Zpracování XML je podporováno řadou nástrojů a programovacích jazyků.
OData	Open Data Protokol (OData) je protokol definovaný společností Microsoft. Verze 1.0 , 2.0 a 3.0 jsou uvolněny pod Microsoft Open Specification Promise. Verze 4.0 je v současné době předmětem standardizace v OASIS. Protokol byl navržen tak, aby umožnil standardní CRUD přístup ke zdroji dat prostřednictvím webové stránky. Je to podobné jako JDBC a ODBC, i když OData není omezen pouze na SQL databázi.
RDF	RDF je jazyk založený na XML sloužící k popisu zdrojů. Jedná se o standardizovaný formát, který umožňuje vyjadřovat popisné informace o WWW zdrojích.

Zdroj: weby výrobců, wikipedia, vlastní tvorba

1. místo: edesky.cz, Marek Aufart

Aplikace přehledně zobrazuje dokumenty vyvěšené na elektronických úředních deskách měst a obcí. Zdrojová data se čerpají z úředních desek jednotlivých krajů, měst a obcí.

2. místo: Dětská hřiště v Praze, Jakub Kuthan, Martin Vašák, Václav Pekárek (FIS VŠE)

Aplikace představuje nejkrásnější dětská hřiště v Praze s popisem atrakcí a zajímavých míst v okolí. K září 2014 obsahuje přehled více než 80 lokalit s asi 130 hřišti. Data jsou získána od příslušných odborů životního prostředí a doplňovány mapovými podklady na bázi Open Street Map.

3. místo: WC kompas, Petr Hřích (občanské sdružení Pacienti IBD)

Mapa veřejných toalet, která umožňuje uživatelům Internetu, počítačů a mobilních zařízení nalézt nejbližší dostupné WC na základě jejich geografické polohy. Web je přizpůsoben zobrazení na chytrém telefonu. Aplikace čerpá data z otevřených dat projektu VozejkMap.cz.

4. místo: Cykloplánovač, Jan Hrnčíř, Michal Jakob, Pavol Žilecký (ČVUT FEL)

Aplikace umožňuje plánování tras pro cyklisty v 6 městech ČR (Praha, Brno, České Budějovice, Plzeň, Hradec Králové, Pardubice) Uživatelé jsou doporučeny 4 plány odpovídající různým potřebám cyklistů (cestovní, klidná, plochá a rychlá trasa). Data pocházejí z Open Street Map a dalších otevřených mapových zdrojů.

5. místo: DATY, Adam Kurzok (Datafusion s.r.o.)

Aplikace sbírá informace o firmách, informuje o jejich aktivitách a upozorňuje na nové obchodní příležitosti nebo rizika. Každý může díky DATY sledovat změny u svých zákazníků, dodavatelů či konkurence, a to přes přehledné webové rozhraní nebo API přímo ve vlastním firemním systému.

6. místo: Road Safety, Petr Schiller (Mfox.cz)

Služba poskytující informace o místech v silniční síti, kde dochází k vyšší koncentraci dopravních nehod se zvěří. V mapové aplikaci lze nastavit vlastní trasu a zjistit riziková místa a ta pak exportovat do vlastních zařízení. Data jsou poskytována z Open Street Map a Ředitelstvím silnic a dálnic ČR.

7. místo: iKatastr.cz, Stanislav Šumbera,

Aplikace umožňuje nahlížení do katastru nemovitostí přes katastrální mapu kliknutím na parcelu-budovu. Aplikace je doplněna také údaji z geografické mapy a informacemi o věcných břemenech, záplavových zónách nebo ekologické zátěži. Služba využívá mapové podklady ČUZK a data z Digitální báze vodohospodářských dat nebo podklady od CENIA.

8. místo: Slepé mapy, Vít Stanislav, Jan Papoušek, (FI MUNI)

Aplikace slouží k procvičování slepých map. Uživatelé se učí určovat polohu jednotlivých států, měst, řek, jezer, pohoří nebo ostrovů. Propracovaný algoritmus volí optimální otázky na základě znalostí daného uživatele. Aplikace je postavena na otevřených mapových podkladech Open Street Map.

9. místo: Datový atlas ČR, Jan Matoušek (Data Mind, s. r. o.)

Aplikace poskytuje široké veřejnosti informace o obyvatelích České republiky, bytovém fondu ze sčítání v roce 2011, dále o struktuře ekonomických subjektů a výsledcích parlamentních voleb v roce 2013. Aplikace využívá otevřená data z Českého statistického úřadu, ARES a část dat z OpenData.cz

10. místo: Koupaliště EU, Michal Ursiny (Fuerte International Ltd.)

Mobilní aplikace pro Android zpřístupňující data o vodních plochách vhodných ke koupání, které jsou evidovány a kontrolovány Evropskou agenturou pro životní prostředí.

Příloha C Ukázky datasetů

Město Děčín, přehled příjmů a výdajů pro rok 2015:

Tabulka č. 1

Rekapitulace					
příjmů, výdajů a financování v roce 2015					
Ukazatel	Skutečnost k 31.12.2013	Schválený rozpočet na rok 2014	Upravený rozpočet k 30. 9. 2014	Rozpočtový výhled na rok 2015	Schválený rozpočet na rok 2015 (v tis. Kč)
Provozní příjmy	720 759	608 400	652 274	672 000	635 016
z toho:					
daňové příjmy	581 883	546 180	552 257	557 000	573 512
nedaňové příjmy	48 305	21 288	25 344	20 000	20 706
dotace	90 571	40 932	74 673	95 000	40 798
Kapitálové příjmy	193 625	2 846	21 317	0	4 000
z toho:					
příjmy z prodeje majetku	6 661	2 846	2 988	0	4 000
investiční dotace	186 964	0	18 329	0	0
PŘÍJMY CELKEM	914 384	611 246	673 591	672 000	639 016
Provozní výdaje	602 632	560 108	631 853	617 000	594 461
Kapitálové výdaje	79 773	90 304	170 960	35 000	87 497
z toho:					
strojní	8 114	10 100	11 109	x	12 145
stavební - rozestavěné	25 967	16 557	31 403	x	32 059
stavební - zahajované	36 587	57 647	120 958	x	32 873
úroky z dlouhodobého úvěru	9 105	6 000	6 000	x	4 000
rezerva	0	0	1 490	x	6 420
VÝDAJE CELKEM	682 405	650 412	802 813	652 000	681 958
FINANCOVÁNÍ, tj. rozdíl mezi příjmy a výdaji	-231 979	39 166	129 222	-20 000	42 942

Návrh rozpočtu	Příjmy	Výdaje	Rozdíl
Provozní rozpočet	635 016	594 461	40 555
Kapitálový rozpočet	4 000	87 497	-83 497
CELKEM	639 016	681 958	-42 942

Obrázek C1. - Přehled příjmů a výdajů Města Děčín pro rok 2015

Zdroj: http://www.mmdecin.cz/dokumenty/cat_view/238-otevrena-data/251-rozpocet-mesta/302-2015

Město Děčín, seznam parcel:

	A	B	C	D	E	F	G
1	KATASTR	PARCELA	CISLO_KMENOVE	CISLO_PODDELENI	VYMER	VYUZITI	DRUH_POZEMKU
2	Březiny u Děčína	-1	1	1	421	zbořeniště	zastavěná plocha a nádvoří
3	Březiny u Děčína	1.1	1	1	1013		zahrada
4	Krásný Studenec	1	1		1047	ostatní komunikace	ostatní plocha
5	Podmokly	1	1		1319		lesní pozemek
6	Podmokly	2.1	2	1	27274	hřbitov, urnový háj	ostatní plocha
7	Podmokly	2.2	2	2	317		zastavěná plocha a nádvoří
8	Krásný Studenec	2.3	2	3	2404	sportoviště a rekreační plocha	ostatní plocha
9	Lesná u Děčína	2.3	2	3	406		lesní pozemek
10	Děčín	2	2		17		zastavěná plocha a nádvoří
11	Chrochvice	3.1	3	1	3082	silnice	ostatní plocha
12	Bělá u Děčína	3.1	3	1	775		trvalý travní porost
13	Chrochvice	3.2	3	2	185	ostatní komunikace	ostatní plocha
14	Lesná u Děčína	3.2	3	2	142	neplodná půda	ostatní plocha
15	Chrochvice	3.3	3	3	82	ostatní komunikace	ostatní plocha
16	Chrochvice	3.4	3	4	28	ostatní komunikace	ostatní plocha
17	Děčín	3	3		287	zeleň	ostatní plocha
18	Folknáře	3	3		2938		trvalý travní porost
19	Hoštice nad Labem	3	3		419		trvalý travní porost
20	Krásný Studenec	3	3		1176	ostatní komunikace	ostatní plocha
21	Křešice u Děčína	3	3		215	ostatní komunikace	ostatní plocha
22	Podmokly	3	3		219		zastavěná plocha a nádvoří
23	Krásný Studenec	4.1	4	1	423		zahrada
24	Chrochvice	4.2	4	2	833	ostatní komunikace	ostatní plocha
25	Krásný Studenec	4.2	4	2	760		zahrada

Obrázek C2. – Seznam parcel v majetku města Děčín

Zdroj: http://www.mmdecin.cz/dokumenty/cat_view/238-otevrena-data/257-majetek-mesta-decin

Město Děčín, evidence veřejných zakázek:

EVIDENCE VEŘEJNÝCH ZAKÁZEK											
Statutární město Děčín			Příloha č. 3 ke směrnici č. 5-6				Tabulka č. 4				
Veřejné zakázky nad 2 000 000 Kč (bez DPH)											
za období 1.1. - 30.6.2014											
číslo odboru/ PO	Odbor MM zajišťující administrativní práce za zadávatele statutární město Děčín/příspěvková organizace zřízená městem	Název vítězného dodavatele	Číslo zakázky	Popis zakázky	Celková částka v Kč vč. DPH	Cena v Kč bez DPH	Předpokládaná hodnota v Kč bez DPH	Rozdíl mezi předpokládanou a konečnou cenou bez DPH	Počet podaných nabídek	Hodnocení	Poznámky
30	Tajemník							-			
31	Odbor provozní a organizační							-			
32	Odbor správních šŕn. a OŽÚ							-			
33	Odbor stavební úřad							-			
35	Odbor ekonomický							-			
36	Odbor rozvoje	INSKY spol. s r.o.	P14V00000013	Bodové výpravy asfaltových komunikací v Děčíně.	475,53	393,00	2 820 000,00		3	v realizaci	rámcová smlouva soutěženo na jednotkovou u cenu
		SalM silnice a mosty Děčín a.s.	P14V00000018	Letmé výpravy asfaltových komunikací v Děčíně.	2 524,06	2 086,00	2 460 000,00		3	v realizaci	rámcová smlouva soutěženo na jednotkovou u cenu
		Viamont DSP a.s.	P12V000000048	Rekonstrukce domova pro seniory, Děčín II, Kamenická 755/195	44 121 489,23	38 358 254,74	59 000 000,00	20 841 745,26	10	v realizaci	
		ACTIVA spol. s r.o.	P13V000000120	Dodávka kancelářských potřeb Revitalizace nádvorek děčínského zámku	2 806 417,44	2 154 064,00	3 800 000,00	1 445 936,00	5	v realizaci	rámcová smlouva
		SalM silnice a mosty Děčín a.s.	P14V000000071	Dodávka elektrické energie pro statutární město Děčín a jím zřízené organizace	6 170 070,30	5 099 231,70	9 729 731	630 499,30	1	nezahájeno	
		Amper Market, a.s.	P13V000000171		23 106 426,20	19 096 220,00	28 000 000,00	8 903 780,00	4	v realizaci	

Obrázek C3. – Evidence veřejných zakázek města Děčín

Zdroj: http://mmdecin.cz/dokumenty/cat_view/136-monitoring-verejnych-zakazek

Město Děčín, benchmarking:

Agenda: 1 - Souhrnné informace

Číslo	Název datového pole	Hodnota 2010	Hodnota 2011	Hodnota 2012	Hodnota 2013
D1.1	Počet obyvatel obce	51 956	51 558	51 488	49 787
D1.1.1	Počet dětí v obci ve věku 3 - 6 let	NA	NA	NA	1 758
D1.2	Počet obyvatel správního obvodu obce s pověřeným obecním úřadem	66 900	66 350	66 278	63 839
D1.3	Počet obyvatel správního obvodu obce s rozšířenou působností	80 508	80 168	80 194	77 818
D1.4	Počet obyvatel správního obvodu obce s působností matričního úřadu	65 840	65 376	65 404	63 766
D1.5	Počet obyvatel správního obvodu obce s působností stavebního úřadu	59 125	58 041	65 404	56 089
D1.6	Počet obyvatel správního obvodu obce pro něž je zajišťováno vyplácení sociálních dávek	80 508	80 168	NA	NA
D1.7	Počet obcí ve správním obvodu obce s pověřeným obecním úřadem	14	14	14	14
D1.8	Počet obcí ve správním obvodu obce s rozšířenou působností	33	33	33	33
D1.9	Počet obcí v působnosti matričního úřadu pro něž obec (vyplňující) zajišťuje působnost matričního úřadu	13	13	13	13
D1.10	Počet matričních úřadů ve správním obvodu	3	3	3	3

Obrázek C4. - Benchmarking města Děčína za období 2010-2013

Zdroj: http://www.mmdecin.cz/dokumenty/cat_view/238-otevrena-data/281-benchmarking-2010-2013

Ukázka smlouvy pomocí výpisu z informačního systému veřejné správy:



MINISTERSTVO VNITRA
ČESKÉ REPUBLIKY

Výpis z informačního systému veřejné správy ve smyslu § 9 odst. 1 zákona č. 365/2000 Sb., o informačních systémech veřejné správy ve znění pozdějších předpisů.

Informační systém veřejné správy a část, ze které se výpis vydává:

Portál veřejné správy > Rejstříky > Registr smluv > Smlouvy

Informace o zápisu:

Rejstřík:	Registr smluv > Smlouvy
Záznam:	128332940
Zveřejňující:	Liberecký kraj
ID datové schránky:	c5kbvkw
Datum zveřejnění:	26.03.2013
Datum a čas poslední změny záznamu:	26.03.2013 09:31:19

Obsah zápisu:

Předmět:	Dodatek č. 8 k Dohodě o úplném znění smlouvy o ZVS ve veřejné linkové osobní dopravě k zabezpečení stanoveného rozsahu ZDO LK pro období od roku 2004 do roku 2013
Partner - název:	COMPAG CZ s.r.o.
Partner - IČ:	62241630
Partner - adresa:	V Lukách 95/IV, Mimoň 47124
Datum vystavení:	22.02.2013
Agenda:	
Číslo dokladu:	OLP/159/2009
Schválil:	Ing. Jan Čáp
Částka bez DPH:	3 600 000,00
Částka vč. DPH:	3 600 000,00

Přílohy záznamu:

Dodatek_8_COMPAG.pdf
159 507 byte, application/pdf
cdce383a724c3d39689ffe8b793a1f578118e58567d1e7820538519aa4fa0841

Historie změn záznamu:

Záznam nemá žádné změny v historii.

Adresa zápisu na Portálu veřejné správy:

<http://portal.gov.cz/portal/obcan/rejstrik/data/10013/2013/03/rec-128332940.html>

Obrázek C5. - Výpis z portálu veřejné správy, smlouva část 1.

Zdroj: <http://portal.gov.cz/portal/rejstrik/data/10013/2013/03/rec-128332940.pdf>

Vysvětlivky:

Informace o zápisu

Údaje v této části jsou generovány automaticky informačním systémem.

Záznam:	Identifikátor záznamu. Hodnota tohoto údaje je shodná s identifikátorem datové zprávy, ve které obdržel správce Portálu veřejné správy údaje ke zveřejnění. Alternativně obsahuje identifikátor záznamu generovaný při ručním vložení záznamu pracovníkem redakce PVS. V případě, že záznam byl označen jako zneplatněný, je hodnota doplněna o text „(ZNEPLATNĚNÝ)“.
Zveřejňující:	Označení subjektu, který záznam zveřejnil.
ID datové schránky:	Identifikátor datové schránky subjektu, který záznam zveřejnil.
Datum zveřejnění:	Hodnotu určuje okamžik dodání datové zprávy do datové schránky „Automat PVS (Ministerstvo vnitra)“, alternativně okamžik vložení pracovníkem redakce PVS.
Datum poslední změny:	Údaj je na výstupu obsažen pouze v případě, že záznam byl po vložení ještě doplněn.
Datum zneplatnění:	Údaj je na výstupu obsažen pouze v případě, že záznam byl zneplatněn.

Obsah zápisu

Údaje v této části byly poskytnuty subjektem, který rejstříkový záznam zveřejnil. Správce Portálu veřejné správy (Ministerstvo vnitra) garantuje, že poskytnuté údaje zveřejnil v nezměněné podobě tak, jak je obdržel. Správce nicméně nemůže garantovat správnost ani úplnost zveřejněných údajů v části „Obsah zápisu“, protože není v jeho moci zjistit ani kontrolovat, zda subjekt, který záznam zveřejnil, poskytl skutečně úplné a autentické údaje.

Přílohy záznamu

Přílohy záznamu jsou uvedeny pouze pro informaci a to v podobě, ve které byly zaslány ke zveřejnění. Není vyžadováno, aby byly zveřejněny v podobě stejnopisu. Pokud záznam obsahuje přílohy, výstup obsahuje seznam příloh v členění (název souboru, velikost v bytech, mime typ souboru, miniatura souboru spočtená dle algoritmu SHA-2).

Historie změn

Pravidla tvorby a zveřejnění rejstříkových záznamů umožňují dodatečné vložení dalších příloh, případně další změny údajů na žádost subjektu, který záznam zveřejnil. Pokud od okamžiku prvotního zveřejnění záznamu došlo k nějaké změně, je o ní veden záznam v této části výpisu.

Obrázek C6. - Výpis z portálu veřejné správy, smlouva část 2.

Zdroj: <http://portal.gov.cz/portal/rejstriky/data/10013/2013/03/rec-128332940.pdf>